

## PCT

**WORLD INTELLECTUAL PROPERTY ORGANIZATION**  
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification 6 :  <b>C12N 9/10, 15/63, 1/21, 1/15, 1/19, C12Q 1/48, C07H 21/04</b></p>	<p><b>A1</b></p>	<p>(11) International Publication Number: <b>WO 99/09145</b></p> <p>(43) International Publication Date: <b>25 February 1999 (25.02.99)</b></p>																																																																																																																																
<p>(21) International Application Number: <b>PCT/US98/16229</b></p> <p>(22) International Filing Date: <b>3 August 1998 (03.08.98)</b></p> <p>(30) Priority Data:  <b>60/055,662</b>      <b>14 August 1997 (14.08.97)</b>      <b>US</b></p> <p>(71) Applicant (for all designated States except US): <b>THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 5th floor, 1111 Franklin Street, Oakland, CA 94607-5200 (US).</b></p> <p>(72) Inventors; and  (75) Inventors/Applicants (for US only): <b>SCHNEEWIND, Olaf [DE/US]; 11050 Strathmore Drive #422, Los Angeles, CA 90024 (US). MAZMANIAN, Sarkis [US/US]; 4401 North Sepulveda Boulevard #310, Sherman Oaks, CA 91403 (US). TON-THAT, Hung [VN/US]; 6021 Piedmont Avenue, Los Angeles, CA 90042 (US).</b></p> <p>(74) Agent: <b>FARBER, Michael, B.; Merchant, Gould, Smith, Edell, Welter &amp; Schmidt, Suite 400, 11150 Santa Monica Boulevard, Los Angeles, CA 90025-3395 (US).</b></p>																																																																																																																																		
<p>(81) Designated States: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</b></p> <p><b>Published</b>  <i>With international search report.</i></p>																																																																																																																																		
<p>(54) Title: <b>IDENTIFICATION OF SORTASE GENE</b></p> <p>(57) Abstract</p> <div style="display: flex;"> <div style="flex: 1; padding-right: 10px;"> <p>The present invention is a substantially purified sortase-transamidase enzyme from Gram-positive bacteria, such as <i>Staphylococcus aureus</i>. The enzyme having a molecular weight of about 41,000 daltons and catalyzing a reaction that covalently cross-links the carboxyl terminus of a protein having a sorting signal to the peptidoglycan of a Gram-positive bacterium, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine, and wherein sorting occurs by cleavage between the fourth and fifth residues of the LPX<sub>3</sub>X<sub>4</sub>G motif. Variants of the enzyme, methods for cloning the gene encoding the enzyme and expressing the cloned gene, and methods of use of the enzyme, including for screening for antibiotics and for display of proteins or peptides on the surfaces of Gram-positive bacteria, are also disclosed.</p> </div> <div style="flex: 2;"> <p style="text-align: right;">Matching Percentage (Total Window: 218, Alignment Window: 228)</p> <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 15%;"><u><b>S.aureus</b></u></td> <td style="width: 5%;">-6</td> <td style="width: 10%;">.....DPK</td> <td style="width: 5%;">L</td> <td style="width: 10%;">K</td> <td style="width: 5%;">EI</td> <td style="width: 10%;">YQ</td> <td style="width: 5%; text-align: right;">43</td> </tr> <tr> <td><u><b>S.pyogenes</b></u></td> <td>1</td> <td>MLQYSQKLPEKFAMSGFLEQRLGHCRLQMAEKGLEALLVTHLTNSYYLTG</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">50</td> </tr> <tr> <td></td> <td>44</td> <td>-----I-----VLE---SCM-----KATNE---I---R-P-----</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">93</td> </tr> <tr> <td></td> <td>51</td> <td>FSGTAATVLITAKRRVLITDSRYTLLAKASVEGFDIIESRTPLKVVAEEL</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">100</td> </tr> <tr> <td></td> <td>94</td> <td>-----G-----MT-----G-----A-----E-----A-----</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">143</td> </tr> <tr> <td></td> <td>101</td> <td>EADQIDCLGFEDQVSFSFYQAMQAEISGITLLAQSGFVEHLRLIKDASEI</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">150</td> </tr> <tr> <td></td> <td>144</td> <td>D-A---IS-----R---NYLE-----SK---</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">193</td> </tr> <tr> <td></td> <td>151</td> <td>DTIAKACSSIDKAFEDALDFIKPGTTTERDLANFLDFMRQYGASGTSFD</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">200</td> </tr> <tr> <td></td> <td>194</td> <td>---GY-----G---K---E---FG---K---SLG---R-G---</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">243</td> </tr> <tr> <td></td> <td>201</td> <td>IIVASGYLSAMPHGRASDKVIONKESLTMDFGCYNNHYVSDMTRTIHIGQ</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">250</td> </tr> <tr> <td></td> <td>244</td> <td>-----IG-LE-----IHE---G-----PM-L---A---R-T---</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">293</td> </tr> <tr> <td></td> <td>251</td> <td>VTDEEREIYALVLAANKALIAKASAGMTYSDFDGIPIQLITEAGYGRFT</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">300</td> </tr> <tr> <td></td> <td>294</td> <td>--IQ-----D-----K---L-Q---VNNCVTVEPGVYIEGLG---GI</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">343</td> </tr> <tr> <td></td> <td>301</td> <td>HGIGHGIGLDIHNPPFGKSEQLLOAGMV---VTDEPGIY---LDNKYGV</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">350</td> </tr> <tr> <td></td> <td>344</td> <td>RIEDDIL-ITENGCOVFTKCTKDLIVLT*.....</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">393</td> </tr> <tr> <td></td> <td>351</td> <td>RIEDD-LVITKTCQVLTLPAPKELIVL*.....</td> <td></td> <td></td> <td></td> <td></td> <td style="text-align: right;">400</td> </tr> </table> <p style="text-align: center;">(SEQ ID No: 2 &amp; 34)</p> </div> </div>			<u><b>S.aureus</b></u>	-6	.....DPK	L	K	EI	YQ	43	<u><b>S.pyogenes</b></u>	1	MLQYSQKLPEKFAMSGFLEQRLGHCRLQMAEKGLEALLVTHLTNSYYLTG					50		44	-----I-----VLE---SCM-----KATNE---I---R-P-----					93		51	FSGTAATVLITAKRRVLITDSRYTLLAKASVEGFDIIESRTPLKVVAEEL					100		94	-----G-----MT-----G-----A-----E-----A-----					143		101	EADQIDCLGFEDQVSFSFYQAMQAEISGITLLAQSGFVEHLRLIKDASEI					150		144	D-A---IS-----R---NYLE-----SK---					193		151	DTIAKACSSIDKAFEDALDFIKPGTTTERDLANFLDFMRQYGASGTSFD					200		194	---GY-----G---K---E---FG---K---SLG---R-G---					243		201	IIVASGYLSAMPHGRASDKVIONKESLTMDFGCYNNHYVSDMTRTIHIGQ					250		244	-----IG-LE-----IHE---G-----PM-L---A---R-T---					293		251	VTDEEREIYALVLAANKALIAKASAGMTYSDFDGIPIQLITEAGYGRFT					300		294	--IQ-----D-----K---L-Q---VNNCVTVEPGVYIEGLG---GI					343		301	HGIGHGIGLDIHNPPFGKSEQLLOAGMV---VTDEPGIY---LDNKYGV					350		344	RIEDDIL-ITENGCOVFTKCTKDLIVLT*.....					393		351	RIEDD-LVITKTCQVLTLPAPKELIVL*.....					400
<u><b>S.aureus</b></u>	-6	.....DPK	L	K	EI	YQ	43																																																																																																																											
<u><b>S.pyogenes</b></u>	1	MLQYSQKLPEKFAMSGFLEQRLGHCRLQMAEKGLEALLVTHLTNSYYLTG					50																																																																																																																											
	44	-----I-----VLE---SCM-----KATNE---I---R-P-----					93																																																																																																																											
	51	FSGTAATVLITAKRRVLITDSRYTLLAKASVEGFDIIESRTPLKVVAEEL					100																																																																																																																											
	94	-----G-----MT-----G-----A-----E-----A-----					143																																																																																																																											
	101	EADQIDCLGFEDQVSFSFYQAMQAEISGITLLAQSGFVEHLRLIKDASEI					150																																																																																																																											
	144	D-A---IS-----R---NYLE-----SK---					193																																																																																																																											
	151	DTIAKACSSIDKAFEDALDFIKPGTTTERDLANFLDFMRQYGASGTSFD					200																																																																																																																											
	194	---GY-----G---K---E---FG---K---SLG---R-G---					243																																																																																																																											
	201	IIVASGYLSAMPHGRASDKVIONKESLTMDFGCYNNHYVSDMTRTIHIGQ					250																																																																																																																											
	244	-----IG-LE-----IHE---G-----PM-L---A---R-T---					293																																																																																																																											
	251	VTDEEREIYALVLAANKALIAKASAGMTYSDFDGIPIQLITEAGYGRFT					300																																																																																																																											
	294	--IQ-----D-----K---L-Q---VNNCVTVEPGVYIEGLG---GI					343																																																																																																																											
	301	HGIGHGIGLDIHNPPFGKSEQLLOAGMV---VTDEPGIY---LDNKYGV					350																																																																																																																											
	344	RIEDDIL-ITENGCOVFTKCTKDLIVLT*.....					393																																																																																																																											
	351	RIEDD-LVITKTCQVLTLPAPKELIVL*.....					400																																																																																																																											

**BEST AVAILABLE COPY**

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

-1-

IDENTIFICATION OF SORTASE GENEGOVERNMENT RIGHTS

This invention was supported by grants from the United States government, namely grants from the National Institutes of Health, NIH-NIAID Grant Nos. AI 33985 and 38897. Accordingly, the government may have certain rights in this invention.

5

BACKGROUND OF THE INVENTION

This invention is directed to an enzyme from Gram-positive bacteria, designated sortase-transamidase, nucleic acid segments encoding the enzyme, and methods of use of the enzyme.

10 Human infections caused by Gram-positive bacteria present a medical challenge due to the dramatic increase in multiple antibiotic resistance strains in recent years. Gram-positive bacteria that can cause serious or fatal infections in humans include *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Pneumococcus*, *Bacillus*, *Actinomyces*, *Mycobacterium*, and *Listeria*, as well as others. Infections  
15 caused by these pathogens are particularly severe and difficult to treat in immunologically compromised patients. These include patients suffering from infection with the Human Immunodeficiency Virus (HIV), the virus that causes AIDS, as well as patients given immune-suppressive agents for treatment of cancer or autoimmune diseases. In particular, infections caused by various *Mycobacterium*  
20 species, including *M. tuberculosis*, *M. bovis*, *M. avium*, and *M. intracellulare*, are frequently the cause of disease in patients with AIDS.

Therefore, it is apparent that new target sites for bacterial chemotherapy are needed if such pathogenic organisms are to be controlled.

25 A unique characteristic of these pathogens and many Gram-positive bacteria is their surface display of proteins anchored to the cell wall. In fact, many of these molecules are known to be involved in essential cellular functions, including pathogenesis in a susceptible host. Thus, a possible disruption in this anchoring process may prove to be an effective treatment against these disease-causing elements.

30 The anchoring of surface molecules to the cell wall in Gram-positive bacteria has been demonstrated to involve a conserved pathway, culminating in recognition of a conserved cleavage/anchoring site by some previously uncharacterized cellular machinery. Molecules whose ultimate location is the cell

-2-

wall must invariably be translocated across the single cellular membrane of these organisms. This is mediated for all cell wall anchored proteins by the well studied secretory pathway, involving cleavage of an amino-terminal signal peptide by a type I signal peptidase. Upon translocation of the molecule out of the cytoplasm, a  
5 mechanism must be present that extracellularly recognizes this protein as a substrate for anchoring. This process has been previously shown to involve the carboxyl-terminally located cell wall sorting signal, consisting of a highly conserved motif such as LPXTG (SEQ ID NO:1), in which X can represent any of the twenty naturally occurring L-amino acids, followed by a series of hydrophobic residues and ultimately  
10 a sequence of positively-charged residues. Thus, once amino-terminally modified and successfully secreted, a polypeptide with this carboxyl-terminal sequence can present itself as a substrate to be processed by the anchoring machinery. At this time, cleavage of the sorting signal after the threonine residue is coupled with covalent linkage of the remainder of the polypeptide to the free amino group of the  
15 pentaglycine crossbridge in the cell wall.

It is this transpeptidation reaction that anchors mature surface proteins so that the peptidoglycan layer, from which point the molecules can serve their biological functions. Therefore, there is a need to isolate and purify the enzyme that catalyzes this reaction. There is also a need to identify the gene encoding such an  
20 enzyme in order that the enzyme can be produced by genetic engineering techniques.

Additionally, there is also a need to develop new methods for displaying proteins or peptides on the surfaces of bacteria. For many purposes, it is desirable to display proteins or peptides on the surfaces of bacteria so that the proteins or peptides are accessible to the surrounding solution, and can, for example, be bound  
25 by a ligand that is bound specifically by the protein or peptide. In particular, the display of proteins on the surface of bacteria is desirable for the preparation of vaccines, the linkage of molecules such as antibiotic molecules or diagnostic reagents to cells, for screening reagents such as monoclonal antibodies, and for the selection of cloned proteins by displaying the cloned proteins, then observing their reaction with  
30 specific reagents such as antibodies. One way of doing this has been with phage display (G.P. Smith, "Filamentous Fusion Phage: Novel Expression Vectors that Display Cloned Antigens on the Virion Surface," *Science* 228:1315-1316 (1985)). However, phage display is limited in its practicality, because it requires that the protein being displayed to be inserted into a coat protein of filamentous phage and  
35 retain its activity while not distorting the conformation of the coat protein, allowing functional virions to be formed. In general, this technique is therefore limited only to small peptide and proteins.

-3-

Therefore, there is a need for a more general method of peptide and protein display.

### SUMMARY

5           The present invention is directed to sortase-transamidase enzymes from Gram-positive bacteria, particularly *Staphylococcus aureus*, and methods for their use, particularly in the areas of drug screening and peptide and protein display.

One aspect of the present invention is a substantially purified sortase-transamidase enzyme from a Gram-positive bacterium, the enzyme catalyzing a  
10    reaction that covalently cross-links the carboxyl terminus of a protein having a sorting signal to the peptidoglycan of a Gram-positive bacterium, the sorting signal having a motif of LPX<sub>3</sub>X<sub>4</sub>G therein, wherein sorting occurs by cleavage between the fourth and fifth residues of the LPX<sub>3</sub>X<sub>4</sub>G motif. Typically, the Gram-positive bacterium is a species selected from the group consisting of *Staphylococcus aureus*, *S. sobrinus*,  
15    *Enterococcus faecalis*, *Streptococcus pyogenes*, and *Listeria monocytogenes*. Preferably, the Gram-positive bacterium is *S. aureus*. The enzyme may be a heterooligomer.

Preferably, the enzyme has at least one subunit with a molecular weight of about 41,000 daltons and the sorting signal further includes: (2) a substantially  
20    hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids  
25    and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine.

Preferably, the enzyme includes therein a subunit whose amino acid sequence is selected from the group consisting of: (1) D-P-K-L-K-E-I-Y-Q-I-V-L-E-S-Q-M-K-A-I-N-E-I-R-P-G-M-T-G-A-E-A-D-A-I-S-R-N-Y-L-E-S-K-G-Y-G-K-E-F-G-H-S-L-G-H-G-I-G-L-E-I-H-E-G-P-M-L-A-R-T-I-Q-D-K-L-Q-V-N-N-C-V-T-V-E-P-G-V-Y-I-E-G-L-G-I-R-I-E-D-D-I-L-I-T-E-N-G-C-Q-V-F-T-K-C-T-K-D-L-I-V-L-T (SEQ ID NO:  
30    2); (2) M-V-K-V-T-D-Y-S-N-S-K-L-G-K-E-I-A-P-E-V-L-S-V-I-A-S-I-A-T-S-E-V-E-G-I-T-G-H-F-A-E-L-K-E-T-N-L-E-K-V-S-R-K-N-L-S-R-D-L-K-I-E-S-K-E-G-I-Y-I-D-V-Y-C-A-L-K-H-G-V-N-I-S-K-T-A-N-K-I-Q-T-S-I-F-N-S-I-S-N-M-T-A-I-E-P-K-Q-I-N-I-H-I-T-Q-I-V-I-E-K (SEQ ID NO: 31) and (3) sequences incorporating one or more conservative amino acid substitutions in SEQ ID NO:2 or SEQ ID NO: 31, wherein  
35

-4-

the conservative amino acid substitutions are any of the following: (1) any of isoleucine, leucine, and valine for any other of these amino acids; (2) aspartic acid for glutamic acid and vice versa; (3) glutamine for asparagine and vice versa; and (4) serine for threonine and vice versa.

- 5 Another aspect of the present invention is a nucleic acid sequence encoding this enzyme. In one alternative, the nucleic acid sequence includes therein a sequence selected from the group consisting of: (1)
- GATCCTAAACTGAAAGAAATATATCAAATAGTACTTGAATCTCAAATGAA  
AGCAATTAATGAGATTAGACCTGGCATGACTGGTGCAGAAGCTGATGCCA  
10 TTTCAAGAACTATTTAGAGTCAAAGGGTATGGAAAAGAATTTGGACAT  
TCACTAGGACATGGTATTGGTTTAGAAATCCATGAAGGGCCAATGCTGGC  
TCGTACGATACAAGATAAACTTCAAGTTAACAACACTGTGTTACAGTAGAAC  
CTGGTGTTTATATAGAAGGTTTGGGCGGTATAAGAATAGAAGATGATATT  
TTAATTACAGAAAATGGTTGTCAAGTCTTTACTAAATGCACAAAAGACCTT  
15 ATAGTTTTAACATAA (SEQ ID NO: 28); (2)
- ATGGTCAAAGTAACTGATTATTCAAATTCAAAATTAGGTAAAGTAGAAAT  
AGCGCCAGAAGTGCTATCTGTTATTGCAAGTATAGCTACTTCGGAAGTCG  
AAGGCATCACTGGCCATTTTGCTGAATTAAGAAACAAATTTAGAAAAA  
GTTAGTCGTAAAAATTTAAGCCGTGATTTAAAAATCGAGAGTAAAGAAGA  
20 TGGCATATATATAGATGTATATTGTGCATTAAACATGGTAATATTTCAA  
AACTGCAAACAAAATTCAAACGTCAATTTTAAATTCAATTTCTAATATGAC  
AGCGATAGAACCCTAAGCAAATTAATATTCACATTACACAAATCGTTATTG  
AAAAGTAA (SEQ ID NO: 30); and (3) a sequence complementary to SEQ ID NO:  
28 or SEQ ID NO: 30. In another alternative, the nucleic acid sequence can include a  
25 sequence hybridizing with SEQ ID NO: 28, SEQ ID NO: 30 or a sequence  
complementary to SEQ ID NO: 28 or SEQ ID NO: 30 with no greater than about a  
15% mismatch under stringent conditions. Preferably, the degree of mismatch is less  
than about 5%; more preferably, the degree of mismatch is less than about 2%.

- Yet another aspect of the present invention is a vector comprising the  
30 nucleic acid sequence of the present invention operatively linked to at least one  
control sequence that controls the expression or regulation of the nucleic acid  
sequence.

Yet another aspect of the present invention is a host cell transfected with a vector of the present invention.

- 35 Another aspect of the present invention is a method for producing a substantially purified sortase-transamidase enzyme. The method comprises the steps of:

-5-

(1) culturing a host cell according to the present invention under conditions in which the host cell expresses the encoded sortase-transamidase enzyme; and

5 (2) purifying the expressed enzyme to produce substantially purified sortase-transamidase enzyme.

Another aspect of the present invention is a method for screening a compound for anti-sortase-transamidase activity. This method is important in providing a way to screen for antibiotics that disrupt the sorting reaction and are likely to be effective in treating infections caused by Gram-positive bacteria.

10 In one alternative, the screening method comprises the steps of:

(1) providing a substantially purified sortase-transamidase enzyme according to the present invention;

(2) performing an assay for sortase-transamidase in the presence and in the absence of the compound; and

15 (3) comparing the activity of the sortase-transamidase enzyme in the presence and in the absence of the compound to screen the compound for sortase-transamidase activity.

In another alternative, the screening method comprises the steps of:

20 (1) providing an active fraction of sortase-transamidase enzyme from a Gram-positive bacterium;

(2) performing an assay for sortase-transamidase in the presence and in the absence of the compound; and

25 (3) comparing the activity of the sortase-transamidase enzyme in the presence and in the absence of the compound to screen the compound for sortase-transamidase activity.

The active fraction of sortase-transamidase activity can be a particulate fraction from *Staphylococcus aureus*.

The assay for sortase-transamidase enzyme can be performed by monitoring the capture of a soluble peptide that is a substrate for the enzyme by its  
30 interaction with an affinity resin. In one alternative, the soluble peptide includes a sequence of at least six histidine residues and the affinity resin contains nickel. In another alternative, the soluble peptide includes the active site of glutathione S-transferase and the affinity resin contains glutathione. In yet another alternative, the soluble peptide includes the active site of streptavidin and the affinity resin contains  
35 biotin. In still another alternative, the soluble peptide includes the active site of maltose binding protein and the affinity resin contains amylose.

Still another aspect of the present invention is an antibody specifically binding the sortase-transamidase enzyme of the present invention.

Yet another aspect of the present invention is a protein molecule comprising a substantially purified sortase-transamidase enzyme according to the present invention extended at its carboxyl-terminus with a sufficient number of  
5 histidine residues to allow specific binding of the protein molecule to a nickel-sepharose column through the histidine residues added at the carboxyl-terminus.

Still another aspect of the present invention is a method for displaying a polypeptide on the surface of a Gram-positive bacterium comprising the steps of:

10 (1) expressing a polypeptide having a sorting signal at its carboxy-terminal end, the sorting signal having: (a) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (b) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (c) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues  
15 being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;

(2) forming a reaction mixture including: (i) the expressed polypeptide; (ii) a substantially purified sortase-transamidase according to the present  
20 invention; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide; and

(3) allowing the sortase-transamidase to catalyze a reaction that cleaves the polypeptide within the LPX<sub>3</sub>X<sub>4</sub> motif of the sorting signal and covalently cross-links the amino-terminal portion of the cleaved polypeptide to the  
25 peptidoglycan to display the polypeptide on the surface of the Gram-positive bacterium.

Another display method according to the present invention comprises:

(1) cloning a nucleic acid segment encoding a chimeric protein into a Gram-positive bacterium to generate a cloned chimeric protein including therein a  
30 carboxyl-terminal sorting signal as described above;

(2) growing the bacterium into which the nucleic acid segment has been cloned to express the cloned chimeric protein to generate a chimeric protein including therein a carboxyl-terminal sorting signal; and

(3) binding the polypeptide covalently to the cell wall by the enzymatic  
35 action of a sortase-transamidase expressed by the Gram-positive bacterium involving cleavage of the chimeric protein within the LPX<sub>3</sub>X<sub>4</sub>G motif so that the polypeptide is



-7-

displayed on the surface of the Gram-positive bacterium in such a way that the polypeptide is accessible to a ligand.

Another aspect of the present invention is a polypeptide displayed on the surface of a Gram-positive bacterium by covalent linkage of an amino-acid  
5 sequence of LPX<sub>3</sub>X<sub>4</sub> derived from cleavage of an LPX<sub>3</sub>X<sub>4</sub>G motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine, the polypeptide being displayed on the surface of the Gram-positive bacterium in such a way that the polypeptide is accessible to a ligand.

10 Another aspect of the present invention is a covalent complex comprising:

- (1) the displayed polypeptide; and
- (2) an antigen or hapten covalently cross-linked to the polypeptide.

Yet another aspect of the present invention is a method for vaccination  
15 of an animal comprising the step of immunizing the animal with the displayed polypeptide to generate an immune response against the displayed polypeptide, or, alternatively, with the covalent complex to generate an immune response against the antigen or the hapten.

Still another aspect of the present invention is a method for screening  
20 for expression of a cloned polypeptide comprising the steps of:

- (1) expressing a cloned polypeptide as a chimeric protein having a sorting signal at its carboxy-terminal end as described above;
- (2) forming a reaction mixture including: (i) the expressed chimeric protein; (ii) a substantially purified sortase-transamidase enzyme according to the  
25 present invention; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide through the sorting signal;
- (3) binding the chimeric protein covalently to the cell wall by the enzymatic action of a sortase-transamidase expressed by the Gram-positive bacterium involving cleavage of the chimeric protein within the LPX<sub>3</sub>X<sub>4</sub>G motif so  
30 that the polypeptide is displayed on the surface of the Gram-positive bacterium in such a way that the polypeptide is accessible to a ligand; and
- (4) reacting the displayed polypeptide with a labeled specific binding partner to screen the chimeric protein for reactivity with the labeled specific binding partner.

35 Still another aspect of the present invention is a method for the diagnosis or treatment of a bacterial infection caused by a Gram-positive bacterium comprising the steps of:

-8-

(1) conjugating an antibiotic or a detection reagent to a protein including therein a carboxyl-terminal sorting signal as described above to produce a conjugate; and

5 (2) introducing the conjugate to an organism infected with a Gram-positive bacterium in order to cause the conjugate to be sorted and covalently cross-linked to the cell walls of the bacterium in order to treat or diagnose the infection.

If an antibiotic is used, typically it is a penicillin, ampicillin, vancomycin, gentamicin, streptomycin, a cephalosporin, amikacin, kanamycin, neomycin, paromomycin, tobramycin, ciprofloxacin, clindamycin, rifampin,  
10 chloramphenicol, norfloxacin, or a derivative of these antibiotics.

Similarly, another aspect of the present invention is a conjugate comprising an antibiotic or a detection reagent covalently conjugated to a protein including therein a carboxyl-terminal sorting signal as described above to produce a conjugate. In still another aspect of the present invention, a composition comprises  
15 the conjugate with a pharmaceutically acceptable carrier.

Another aspect of the present invention is a substantially purified protein having at least about 50% match with best alignment with the amino acid sequences of at least one of the putative *Bacillus* peptidase (SEQ ID NO: 3), the aminopeptidase P of *Lactococcus lactis* (SEQ ID NO: 4), or the proline dipeptidase of  
20 *Lactobacillus delbrueckii lactis* (SEQ ID NO: 5) and having sortase-transamidase activity. Preferably, the match is at least about 60% in best alignment; more preferably, the match is at least about 70% in best alignment.

Another aspect of the present invention is a substantially purified protein having sortase-transamidase activity and a hydrophobicity profile of at least  
25 one subunit of the protein that, determined as the mean absolute value of the hydrophobicity difference per residue, differs from the hydrophobicity profile of a putative *Bacillus* peptidase (SEQ ID NO: 3) by no more than about 2 units on the hydrophobicity scale. Preferably, the difference is not more than about 1 unit; most preferably, it is not more than about 0.5 units.

30 Another aspect of the present invention is a substantially purified protein having sortase-transamidase activity and a hydrophobicity profile of at least one subunit of the protein that, determined as the mean absolute value of the hydrophobicity difference per residue, differs from the hydrophobicity profile of the sequence of SEQ ID NO: 31 by no more than about 2 units on the hydrophobicity  
35 scale.

### BRIEF DESCRIPTION OF THE DRAWINGS

These and other features, aspects, and advantages of the present invention will become better understood with reference to the following description and accompanying drawings where:

Figure 1 is a diagram showing the substantial homology of the amino acid sequence of the sortase-transamidase enzyme of *Staphylococcus aureus* to an open reading frame in the genome of *Streptococcus pyogenes* (SEQ ID NO: 2 & 34);

Figure 2 is a diagram showing a greater homology of the amino acid sequence of the sortase-transamidase enzyme of *Staphylococcus aureus* to the carboxyl-terminal portion of the open reading frame in the genome of *Streptococcus pyogenes* (SEQ ID NO: 2 & 34);

Figure 3 is the DNA sequence of the *S. pyogenes* open reading frame (SEQ ID NO: 33 & 34);

Figure 4 (SEQ ID NO: 34) is the entire amino acid sequence of the protein translated from the entire *S. pyogenes* open reading frame;

Figure 5 (SEQ ID NO: 3) is the amino acid sequence of a putative *Bacillus* peptidase in the GCVT-SPOIIIAA intergenic region;

Figure 6 is the hydrophobicity profile of the protein whose amino acid sequence is shown in Figure 5 (SEQ ID NO: 3);

Figure 7 (SEQ ID NO: 4) is the amino acid sequence of the aminopeptidase P of *Lactococcus lactis*;

Figure 8 (SEQ ID NO: 5) is the amino acid sequence of the proline dipeptidase of *Lactobacillus delbrueckii lactis*;

Figure 9 is a diagram of the activity of the sortase-transamidase enzyme of the present invention;

Figure 10 (SEQ ID NOS: 28 & 29) is a partial DNA sequence of the gene for one of the subunits of the sortase-transamidase enzyme of *S. aureus*;

Figure 11 (SEQ ID NO: 2) is the partial carboxyl-terminal amino acid sequence translated from the DNA sequence of Figure 10 (SEQ ID NOS: 28 & 29);

Figure 12 (SEQ ID NOS: 30 & 31) is a partial DNA sequence of the gene for a second of the subunits of the sortase-transamidase enzyme of *S. aureus*; and

Figure 13 is the hydrophobicity profile of the protein translated from the DNA sequence of Figure 12 (SEQ ID NOS: 30 & 31).

-10-

### DEFINITIONS

As used herein, the terms defined below have the following meanings unless otherwise indicated:

- 5 "Nucleic Acid Sequence": the term "nucleic acid sequence" includes both DNA and RNA unless otherwise specified, and, unless otherwise specified, includes both double-stranded and single-stranded nucleic acids. Also included are hybrids such as DNA-RNA hybrids. In particular, a reference to DNA includes RNA that has either the equivalent base sequence except for the substitution of uracil and RNA for thymine in DNA, or has a complementary base sequence except for the
- 10 substitution of uracil for thymine, complementarity being determined according to the Watson-Crick base pairing rules. Reference to nucleic acid sequences can also include modified bases as long as the modifications do not significantly interfere either with binding of a ligand such as a protein by the nucleic acid or with Watson-Crick base pairing.
- 15 "Antibody": as used herein the term "antibody" includes both intact antibody molecules of the appropriate specificity, and antibody fragments (including Fab, F(ab'), Fv, and F(ab')<sub>2</sub>), as well as chemically modified intact antibody molecules and antibody fragments, including hybrid antibodies assembled by in vitro reassociation of subunits. Also included are single-chain antibody molecules
- 20 generally denoted by the term sFv and humanized antibodies in which some or all of the originally non-human constant regions are replaced with constant regions originally derived from human antibody sequences. Both polyclonal and monoclonal antibodies are included unless otherwise specified. Additionally included are modified antibodies or antibodies conjugated to labels or other molecules that do not
- 25 block or alter the binding capacity of the antibody.

### DESCRIPTION

A substantially purified sortase-transamidase enzyme from Gram-positive bacteria, particularly *Staphylococcus aureus*.

- 30 The properties of this enzyme make it a logical target for antibiotic action. This enzyme also catalyzes covalent crosslinkage of proteins to the peptidoglycan of Gram-positive bacteria.

#### I. THE SORTASE-TRANSAMIDASE ENZYME

- 35 One aspect of the invention is a substantially purified sortase-transamidase enzyme from a Gram-positive bacterium. As used herein, the term "substantially purified" means having a specific activity of at least tenfold greater than

-11-

the sortase-transamidase activity present in a crude extract, lysate, or other state from which proteins have not been removed and also in substantial isolation from proteins found in association with sortase-transamidase in the cell.

One subunit of the enzyme has a molecular weight of about 41,000 daltons. The enzyme catalyzes a reaction that covalently crosslinks the carboxyl-terminus of a protein having a sorting signal to the peptidoglycan of the Gram-positive bacterium. The sorting signal has: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif. In this sorting signal, X<sub>3</sub> can be any of the twenty naturally-occurring L-amino acids. X<sub>4</sub> can be alanine, serine, or threonine. Preferably, X<sub>4</sub> is threonine.

The sortase-transamidase is believed to occur in all Gram-positive bacteria. In particular, the enzyme exists in *Mycobacterium*, *Nocardia*, *Actinomyces*, *Staphylococcus*, *Streptococcus*, *Listeria*, *Enterococcus*, and *Pneumococcus*. Specifically, the enzyme exists in the following species: *Staphylococcus aureus*, *S. sobrinus*, *Enterococcus faecalis*, *Streptococcus pyogenes*, and *Listeria monocytogenes*.

Preferably the enzyme is isolated from *Staphylococcus aureus*.

#### A. Amino Acid Sequence

The sortase-transamidase of the present invention includes therein an amino acid sequence, in one subunit of the enzyme, of D-P-K-L-K-E-I-Y-Q-I-V-L-E-S-Q-M-K-A-I-N-E-I-R-P-G-M-T-G-A-E-A-D-A-I-S-R-N-Y-L-E-S-K-G-Y-G-K-E-F-G-H-S-L-G-H-G-I-G-L-E-I-H-E-G-P-M-L-A-R-P-I-Q-D-K-L-Q-V-N-N-C-V-T-V-E-P-G-V-Y-I-E-G-L-G-G-I-R-I-E-D-D-I-L-I-T-E-N-G-C-Q-V-F-T-K-C-T-K-D-L-I-V-L-T (SEQ ID NO:2). This sequence is at the carboxyl-terminal end of the subunit of the enzyme.

The sortase-transamidase of the present invention also includes therein an amino acid sequence, in a second subunit of the enzyme, of M-V-K-V-T-D-Y-S-N-S-K-L-G-K-E-I-A-P-E-V-L-S-V-I-A-S-I-A-T-S-E-V-E-G-I-T-G-H-F-A-E-L-K-E-T-N-L-E-K-V-S-R-K-N-L-S-R-D-L-K-I-E-S-K-E-G-I-Y-I-D-V-Y-C-A-L-K-H-G-V-N-I-S-K-T-A-N-K-I-Q-T-S-I-F-N-S-I-S-N-M-T-A-I-E-P-K-Q-I-N-I-H-I-T-Q-I-V-I-E-K (SEQ ID NO: 31).

Also within the scope of the present invention are substantially purified protein molecules that are mutants of the sequence of SEQ ID NO:2 or SEQ ID NO: 31 that preserve the sortase-transamidase activity. In particular, the conservative amino acid substitutions can be any of the following: (1) any of isoleucine, leucine, and valine for any other of these amino acids; (2) aspartic acid for glutamic acid and vice versa; (3) glutamine for asparagine and vice versa; and (4) serine for threonine and vice versa.

Other substitutions can also be considered conservative, depending upon the environment of the particular amino acid. For example, glycine (G) and alanine (A) can frequently be interchangeable, as can be alanine and valine (V). Methionine (M), which is relatively hydrophobic, can frequently be interchanged with leucine and isoleucine, and sometimes with valine. Lysine (K) and arginine (R) are frequently interchangeable in locations in which the significant feature of the amino acid residue is its charge and the different pK's of these two amino acid residues or their different sizes are not significant. Still other changes can be considered "conservative" in particular environments. For example, if an amino acid on the surface of a protein is not involved in a hydrogen bond or salt bridge interaction with another molecule, such as another protein subunit or a ligand bound by the protein, negatively charged amino acids such as glutamic acid and aspartic acid can be substituted for by positively charged amino acids such as lysine or arginine and vice versa. Histidine (H), which is more weakly basic than arginine or lysine, and is partially charged at neutral pH, can sometimes be substituted for these more basic amino acids. Additionally, the amides glutamine (Q) and asparagine (N) can sometimes be substituted for their carboxylic acid homologues, glutamic acid and aspartic acid.

The sortase-transamidase from *Staphylococcus aureus* has substantial homology the amino acid sequence of the first subunit, that of SEQ ID NO: 2, with an open reading frame in the genome of *Streptococcus pyogenes*, particularly in the amino-terminal region. There is about a 22% match with best alignment over the entire sequenced region of the *S. pyogenes* open reading frame, and about a 47% match with best alignment over the carboxyl-terminal region of the *S. pyogenes* open reading frame. These matches are shown in Figures 1-2. The DNA sequence of the entire *S. pyogenes* open reading frame is shown in Figure 3 (SEQ ID NO: 33 & 34). The protein translated from the entire *S. pyogenes* open reading frame has a molecular weight of about 40,851.43 daltons; its sequence is shown in Figure 4 (SEQ ID NO: 34). Therefore, another aspect of the present invention is a substantially purified protein molecule that has at least one subunit of about 40,000 to about 41,000 daltons

in molecular weight, the subunit having at least a 20% match with best alignment with the *S. pyogenes* open reading frame of Figure 2 and that has sortase-transamidase activity. Preferably, the subunit has at least a 30% match with best alignment; more preferably, at least 50% match with best alignment.

5           As shown below in the Example, the first 364 bases of a nucleic acid segment that complements a temperature-sensitive mutation in the *S. aureus* sortase-transamidase, designated the SM-317 complementing gene insert, has been identified as encoding a protein sequence that is a homologue of a putative *Bacillus* peptidase in the GCVT-SPOIIIAA intergenic region (GenBank Accession No. 1731048; Y. Kobayashi et al.). The sequence of this putative peptidase is shown in Figure 5 (SEQ ID NO:3) and its hydrophobicity profile is shown in Figure 6. The hydrophobicity is calculated according to the method of J. Kyte & R.F. Doolittle, "A Simple Method for Displaying the Hydropathic Character of a Protein," *J. Mol. Biol.* 157: 105-132 (1982). As used herein, the term "hydrophobicity" is the hydrophobicity as calculated  
10           in Kite & Doolittle, supra.

15           To a lesser degree of homology, the protein sequence encoded by this complementing gene insert is homologous to aminopeptidase P of *Lactococcus lactis* (GenBank Accession No. 1915907; J. Matos). The amino acid sequence of this aminopeptidase is shown in Figure 7 (SEQ ID NO: 4). To a still lesser degree of  
20           homology, the protein sequence encoded by this complementing gene insert is homologous to the proline dipeptidase of *Lactobacillus delbrueckii lactis* (GenBank Accession No. 1172066; K. Stucky et al., "Cloning and DNA Sequence Analysis of pepQ, a Prolidase Gene from *Lactobacillus delbrueckii* subsp. *lactis* and Partial Characterization of Its Product," *Mol. Gen. Genet.* 247: 494-500 (1995)). The amino  
25           acid sequence of this proline dipeptidase is shown in Figure 8 (SEQ ID NO:5).

          Because of the relatedness of these proteins, another aspect of the present invention is a substantially purified protein having at least one subunit with at least about 50% match with best alignment with the amino acid sequences of at least one of the putative *Bacillus* peptidase (SEQ ID NO: 3), the aminopeptidase P of  
30           *Lactococcus lactis* (SEQ ID NO: 4), or the proline dipeptidase of *Lactobacillus delbrueckii lactis* (SEQ ID NO: 5) and having sortase-transamidase activity. Preferably, the at least one subunit of the protein has at least about 60% match with best alignment with at least one of these sequence; more preferably, the at least one subunit of the protein has at least about 70% match with best alignment with at least  
35           one of these sequences.

          Because the hydrophobicity of a protein is a sensitive measure of protein structure, another aspect of the invention is a substantially purified protein

-14-

having sortase-transamidase activity and a hydrophobicity profile for at least one subunit of the protein that, determined as the mean absolute value of the hydrophobicity difference per residue, differs from the hydrophobicity profile of the putative *Bacillus* peptidase by no more than about 2 units on the hydrophobicity scale of Kyte & Doolittle, supra. Preferably, the difference is no greater than about 1 unit; more preferably, the difference is no greater than about 0.5 units.

The sortase-transamidase is a cysteine protease.

#### B. Activity of the Sortase-Transamidase

The activity of the sortase-transamidase enzyme of the present invention is shown, in general, in Figure 9. The enzyme first cleaves a polypeptide having a sorting signal within the LPX<sub>3</sub>X<sub>4</sub>G motif. Cleavage occurs after residue X<sub>4</sub>, normally a threonine; as indicated above, this residue can also be a serine or alanine residue. This residue forms a covalent intermediate with the sortase. The next step is the transamidation reaction that transfers the cleaved carboxyl terminus of the protein to be sorted to the -NH<sub>2</sub> of the pentaglycine crossbridge within the peptidoglycan precursor. The peptidoglycan precursor is then incorporated into the cell wall by a transglycosylase reaction with the release of undecaprenyl phosphate. The mature anchored polypeptide chains are thus linked to the pentaglycine cross bridge in the cell wall which is tethered to the ε-amino side chain of an unsubstituted cell wall tetrapeptide. A carboxypeptidase may cleave a D-Ala-D-Ala bond of the pentapeptide structure to yield the final branched anchor peptide in the staphylococcal cell wall.

The sorting signal has: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region.

In the motif, X<sub>3</sub> can be any of the 20 naturally-occurring L-amino acids. X<sub>4</sub> can be any of threonine, serine, or alanine. Preferably, X<sub>4</sub> is threonine (O. Schneewind et al., "Cell Wall Sorting Signals in Surface Proteins of Gram-Positive Bacteria," EMBO J. 12:4803-4811 (1993)).

Preferably, the substantially hydrophobic domain carboxyl to the motif includes no more than about 7 charged residues or residues with polar side chains. For the purposes of this specification, these residues include the following: aspartic acid, glutamic acid, lysine, and arginine as charged residues, and serine, threonine, glutamine, and asparagine as polar but uncharged residues. Preferably, the sequence includes no more than three charged residues.



Representative sequences suitable for sorting signals for use with the sortase-transamidase of the present invention include, but are not limited to the following: E-E-N-P-F-I-G-T-T-V-F-G-G-L-S-L-A-L-G-A-A-L-L-A-G (SEQ ID NO: 6), the hydrophobic domain of the staphylococcal proteinase (SPA) sorting signal from *Staphylococcus aureus*; (2) G-E-E-S-T-N-K-G-M-L-F-G-G-L-F-S-I-L-G-L-A-L-L (SEQ ID NO:7), the SNBP signal of *S. aureus*; (3) D-S-S-N-A-Y-L-P-L-L-G-L-V-S-L-T-A-G-F-S-L-L-G-L (SEQ ID NO: 8), the SPAA signal of *S. sobrinus*, (4) E-K-Q-N-V-L-L-T-V-V-G-S-L-A-A-M-L-G-L-A-G-L-G-F (SEQ ID NO:9), the PRGB signal of *Enterococcus faecalis*, (5) S-I-G-T-Y-L-F-K-I-G-S-A-A-M-I-G-A-I-G-I-Y-I-V (SEQ ID NO:10), the TEE signal of *Streptococcus pyogenes*, and (6) D-S-D-N-A-L-Y-L-L-L-G-L-L-A-V-G-T-A-M-A-L-T (SEQ ID NO:11), the INLA signal of *Listeria monocytogenes*. Other hydrophobic domains can be used as part of the sorting signal.

The third portion of the sorting signal is a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain. At least one of the two positively charged residues is arginine. The charged tail can also contain other charged amino acids, such as lysine. Preferably, the charged tail region includes two or more arginine residues. The two positively charged residues are located at residues 31-33 from the motif. Preferably, the two arginine residues are either in succession or are separated by no more than one intervening amino acid. Preferably, the charged tail is at least five amino acids long, although four is possible. Among the charged tails that can be used are the following: (1) R-R-R-E-L (SEQ ID NO:12), from the SPA signal of *S. aureus*; (2) R-R-N-K-K-N-H-K-A (SEQ ID NO:13), from the SNBP signal of *S. aureus*; (3) R-R-K-Q-D (SEQ ID NO:14), from the SPAA signal of *S. sobrinus*; (4) K-R-R-K-E-T-K (SEQ ID NO:15), from the PRGB signal of *E. faecalis*; (5) K-R-R-K-A (SEQ ID NO:16), from the TEE signal of *S. pyogenes*; (6), K-R-R-H-V-A-K-H (SEQ ID NO:17), from the FIM sorting signal of *Actinomyces viscosus*, and (7) K-R-R-K-S (SEQ ID NO:18), from the BAC sorting signal of *Streptococcus agalactiae*; (8) K-R-K-E-E-N (SEQ ID NO:19), from the EMM signal of *Streptococcus pyogenes*.

Also usable as the charged tail portion of the sorting signal are the following sequences produced by mutagenesis from the SPA signal of *S. aureus*. These include R-R-R-E-S (SEQ ID NO: 20), R-R-R-S-L (SEQ ID NO: 21), R-R-S-E-L (SEQ ID NO: 22), R-S-R-E-L (SEQ ID NO: 23) and S-R-R-E-L (SEQ ID NO: 24). Other charged tails that are usable as part of the sorting signal can be derived from a polyserine tail, itself inactive, by replacement of one or more of the serine residues with the basic amino acid arginine. These include R-R-S-S-S (SEQ

-16-

ID NO: 25), R-S-R-S-S (SEQ ID NO:26), and S-R-R-S-S (SEQ ID NO:27).

Other sorting signals can also be used.

## II. THE GENE ENCODING THE SORTASE-TRANSAMIDASE ENZYME

### 5 A. Isolation of the Sortase-Transamidase Enzyme Gene

The gene for the sortase-transamidase enzyme in *Staphylococcus aureus* has been isolated. The isolation process is described in detail in the Example below; in general, this process comprises: (1) the generation of temperature-sensitive mutants through chemical mutagenesis, such as with the DNA modifying agent N-methyl-N-nitro-N-nitrosoguanidine; (2) Screening for temperature-sensitive  
10 mutants; (3) screening the temperature-sensitive mutants for a block in protein sorting by the use of a construct harboring the staphylococcal enterotoxin B (SEB) gene fused to the cell wall sorting signal of staphylococcal Protein A (SPA), to locate mutants that accumulate a precursor molecule formed by cleavage of an amino-terminal signal  
15 peptide but that is not then processed by cleavage of the carboxyl-terminal sorting signal; (4) generation of a *S. aureus* chromosomal library and complementation of the temperature-sensitive sorting defect; and (5) sequencing and characterization of the *S. aureus* complementing determinants.

### 20 B. Sequence of the Sortase-Transamidase Gene

The above procedure yielded a partial sequence for one of the subunits of the sortase-transamidase including the carboxyl-terminal portion of the gene for the first subunit. This sequence is

GATCCTAAACTGAAAGAAATATATCAAATAGTACTTGAATCTCAAATGAA  
25 AGCAATTAATGAGATTAGACCTGGCATGACTGGTGCAGAAGCTGATGCCA  
TTTCAAGAACTATTTAGAGTCAAAGGGTATGGAAAAGAATTTGGACAT  
TCACTAGGACATGGTATTGGTTTAGAAATCCATGAAGGGCCCAATGCTGGC  
TCGTACGATACAAGATAAACTTCAAGTTAACAACACTGTGTTACAGTAGAAC  
CTGGTGTTTATAGAAGGTTTGGGCGGTATAAGAATAGAAGATGATATTTT  
30 AATTACAGAAAATGGTTGTCAAGTCTTTACTAAATGCACAAAAGACCTTA  
TAGTTTAAACATAA (SEQ ID NO:28 & 29). The last three nucleotides, TAA, of this sequence are the stop codon.

The above procedure further yielded a sequence for a second subunit of  
ATGGTCAAAGTAACTGATTATTCAAATTCAAAATTAGGTAAAGTAGAAAT  
35 AGCGCCAGAAGTGCTATCTGTTATTGCAAGTATAGCTACTTCGGAAGTCG  
AAGGCATCACTGGCCATTTTGCTGAATTAAGAAACAAATTTAGAAAAA  
GTTAGTCGTAAAAATTTAAGCCGTGATTTAAAAATCGAGAGTAAAGAAGA

-17-

TGGCATATATATAGATGTATATTGTGCATTAAAACATGGTAATATTTCAAA  
AACTGCAAACAAAATTCAAACGTCAATTTTAAATTCAATTTCTAATATGAC  
AGCGATAGAACCTAAGCAAATTAATATTACATTACACAAATCGTTATTG  
AAAAGTAA (SEQ ID NO: 30 & 31) The last three nucleotides of this sequence,

5 TAA, are the stop codon.

Accordingly, within the scope of the present invention is a nucleic acid sequence encoding a substantially purified sortase-transamidase enzyme from a Gram-positive bacterium. The enzyme encoded has at least one subunit with a molecular weight of about 41,000 daltons and catalyzes a reaction that covalently  
10 cross-links the carboxyl-terminus of a protein having the sorting signal described above to the peptidoglycan of a gram-positive bacterium. The nucleic acid sequence includes therein the sequence of SEQ ID NO: 28 or a sequence complementary to SEQ ID NO: 28, or the sequence of SEQ ID NO: 30 or a sequence complementary to SEQ ID NO: 30.

Also included within the present invention is a nucleic acid sequence encoding a substantially purified sortase-transamidase enzyme from a Gram-positive bacterium with at least one subunit with a molecular weight of about 41,000 daltons, where the enzyme catalyzes the cross-linking reaction where the nucleic acid sequence hybridizes with at least one of: (1) the sequence of SEQ ID NO: 28; (2) a  
20 sequence complementary to SEQ ID NO: 28; (3) the sequence of SEQ ID NO: 30; or (4) a sequence complementary to SEQ ID NO: 30 with no greater than about a 15% mismatch under stringent conditions. Preferably, the degree of mismatch is no greater than about 5%; most preferably the mismatch is no greater than about 2%.

Also within the present invention is a nucleic acid sequence encoding a  
25 substantially purified sortase-transamidase enzyme from a Gram-positive bacterium where the enzyme has at least one subunit with a molecular weight of about 41,000 daltons and catalyzes the cross-linking reaction described above involving the sorting signal, where the enzyme includes therein an amino acid sequence selected from the group consisting of: (1) D-P-K-L-K-E-I-Y-Q-I-V-L-E-S-Q-M-K-A-I-N-E-  
30 I-R-P-D-M-T-G-A-E-A-D-A-I-S-R-N-Y-L-E-S-K-G-Y-G-K-E-F-G-H-S-L-G-H-G-I-G-L-E-I-H-E-G-P-M-L-A-R-T-I-Q-D-K-L-Q-V-N-N-C-V-T-V-E-P-G-V-Y-I-E-G-L-G-I-R-I-E-D-D-I-L-I-T-E-N-G-C-Q-V-F-T-K-C-T-K-D-L-I-V-L-T (SEQ ID NO:2); (2) M-V-K-V-T-D-Y-S-N-S-K-L-G-K-E-I-A-P-E-V-L-S-V-I-A-S-I-A-T-S-E-V-E-G-I-T-  
35 G-H-F-A-E-L-K-E-T-N-L-E-K-V-S-R-K-N-L-S-R-D-L-K-I-E-S-K-E-G-I-Y-I-D-V-Y-C-A-L-K-H-G-V-N-I-S-K-T-A-N-K-I-Q-T-S-I-F-N-S-I-S-N-M-T-A-I-E-P-K-Q-I-N-I-H-I-T-Q-I-V-I-E-K (SEQ ID NO:

31); and (3) sequences incorporating one or more conservative amino acid substitutions in SEQ ID NO:2 or SEQ ID NO: 31 wherein the conservative amino acid substitutions are any of the following: (1) any of isoleucine, leucine and valine for any other of these amino acids; (2) aspartic acid for glutamic acid and vice versa; (3) glutamine for asparagine and vice versa; and (4) serine for threonine and vice versa. Alternative nucleic acid sequences can be determined using the standard genetic code; the alternative codons are readily determinable for each amino acid in this sequence.

Construction of nucleic acid sequences according to the present invention can be accomplished by techniques well known in the art, including solid-phase nucleotide synthesis, the polymerase chain reaction (PCR) technique, reverse transcription of DNA from RNA, the use of DNA polymerases and ligases, and other techniques. If an amino acid sequence is known, the corresponding nucleic acid sequence can be constructed according to the genetic code.

#### 15 C. Vectors and Host Cells Transformed with Vectors

Another aspect of the invention is a vector comprising a nucleic acid sequence according to the present invention operatively linked to at least one control sequence that controls the expression or regulation of the nucleic acid sequence. Such control sequences are well known in the art and include operators, promoters, enhancers, promoter-proximal elements and replication origins. The techniques of vector construction, including cloning, ligation, gap-filling, the use of the polymerase chain reaction (PCR) procedure, solid-state oligonucleotide synthesis, and other techniques, are all well known in the art and need not be described further here.

Another aspect of the present invention is a host cell transfected with a vector according to the present invention. Among the host cells that can be used are gram-positive bacteria such as *Staphylococcus aureus*.

Transfection, also known as transformation, is done using standard techniques appropriate to the host cell used, particularly *Staphylococcus aureus*. Such techniques are described, for example, in R.P. Novick, "Genetic Systems in *Staphylococci*," Meth. Enzymol. 204: 587-636 (1991), as well as in O. Schneewind et al., "Sorting of Protein A to the Staphylococcal Cell Wall," Cell 70: 267-281 (1992).

### III. SORTASE-TRANSAMIDASE AS A TARGET FOR ANTIBIOTIC ACTION

#### A. A Site for Antibiotic Action

The reaction carried out by the sortase-transamidase of the present invention presents a possible target for a new class of antibiotics to combat medically relevant infections caused by numerous gram-positive organisms. Because this is a

novel site of antibiotic action, these antibiotics have the advantage that resistance by the bacterium has not had a chance to develop.

Such antibiotics can include compounds with structures that mimic the cleavage site, such as compounds with a structure similar to methyl

5 methanethiosulfonate or, more generally, alkyl methanethiosulfonates. The sortase-transamidase of the present invention is believed to be a cysteine protease. Other antibiotics that may inhibit the activity of the sortase-transamidase in the present invention include inhibitors that would be specific for cysteine-modification in a  $\beta$ -lactam framework. These inhibitors would have active moieties that would form

10 mixed disulfides with the cysteine sulfhydryl. These active moieties could be derivatives of methanethiosulfonate, such as methanethiosulfonate ethylammonium, methanethiosulfonate ethyltrimethylammonium, or methanethiosulfonate ethylsulfonate (J.A. Javitch et al., "Mapping the Binding Site Crevice of the Dopamine D2 Receptor by the Substituted-Cysteine Accessibility Method," Neuron,

15 14: 825-831 (1995); M.H. Akabas & A. Karlin, "Identification of Acetylcholine Receptor Channel-Lining Residues in the M1 Segment of the  $\alpha$ -Subunit," Biochemistry 34: 12496-12500 (1995)). Similar reagents, such as alkyl alkanethiosulfonates, i.e., methyl methanethiosulfonate, or alkoxycarbonylalkyl disulfides, have been described (D.J. Smith et al., "Simple Alkanethiol Groups for

20 Temporary Blocking of Sulfhydryl Groups of Enzymes," Biochemistry 14: 766-771 (1975); W.N. Valentine & D.E. Paglia, "Effect of Chemical Modification of Sulfhydryl Groups of Human Erythrocyte Enzymes," Am. J. Hematol. 11: 111-124 (1981)). Other useful inhibitors involve derivatives of 2-trifluoroacetylaminobenzene sulfonyl fluoride (J.C. Powers, "Proteolytic Enzymes and Their Active-Site-Specific

25 Inhibitors: Role in the Treatment of Disease," in Modification of Proteins), in a  $\beta$ -lactam framework, peptidyl aldehydes and nitriles (E. Dufour et al., "Peptide Aldehydes and Nitriles as Transition State Analog Inhibitors of Cysteine Proteases," Biochemistry 34: 9136-9143 (1995); J. O. Westerik & R. Wolfenden, "Aldehydes as Inhibitors of Papain," J. Biol. Chem. 247: 8195-8197 (1972)), peptidyl diazomethyl

30 ketones (L. Björck et al., "Bacterial Growth Blocked by a Synthetic Peptide Based on the Structure of a Human Proteinase Inhibitor," Nature 337: 385-386 (1989)), peptidyl phosphoramidates (P.A. Bartlett & C.K. Marlowe, "Phosphoramidates as Transition-State Analogue Inhibitors of Thermolysin," Biochemistry 22: 4618-4624 (1983)), phosphonate monoesters such as derivatives or analogues of m-

35 carboxyphenyl phenylacetamidomethylphosphonate (R.F. Pratt, "Inhibition of a Class C  $\beta$ -Lactamase by a Specific Phosphonate Monoester," Science 246: 917-919 (1989)), maleimides and their derivatives, including derivatives of such bifunctional

maleimides as o-phenylenebismaleimide, p-phenylenebismaleimide, m-phenylenebismaleimide, 2,3-naphthalenebismaleimide, 1,5-naphthalenebismaleimide, and azophenylbismaleimide, as well as monofunctional maleimides and their derivatives (J.V. Moroney et al., "The Distance Between Thiol  
5 Groups in the  $\gamma$  Subunit of Coupling Factor 1 Influences the Proton Permeability of Thylakoid Membranes," J. Bioenerget. Biomembr. 14: 347-359 (1982)), peptidyl halomethyl ketones (chloromethyl or fluoromethyl ketones), peptidyl sulfonium salts, peptidyl acyloxymethyl ketones, derivatives and analogues of epoxides, such as E-64 (N-[N-(L-trans-carboxyoxiran-2-carbonyl)-L-leucylagmatine), E-64c (a  
10 derivative of E-64 in which the agmatine moiety is replaced by an isoamylamine moiety), E-64c ethyl ester, Ep-459 (an analogue of E-64 in which the agmatine moiety is replaced by a 1,4-diaminopropyl moiety), Ep-479 (an analogue of E-64 in which the agmatine moiety is replaced by a 1,7-diheptylamino moiety), Ep-460 (a derivative of Ep-459 in which the terminal amino group is substituted with a Z  
15 (benzyloxycarbonyl) group), Ep-174 (a derivative of E-64 in which the agmatine moiety is removed, so that the molecule has a free carboxyl residue from the leucine moiety), Ep-475 (an analogue of E-64 in which the agmatine moiety is replaced with a  $\text{NH}_2-(\text{CH}_2)_2-\text{CH}-(\text{CH}_3)_2$  moiety), or Ep-420 (a derivative of E-64 in which the hydroxyl group is benzoylated, forming an ester, and the leucylagmatine moiety is  
20 replaced with isoleucyl-O-methyltyrosine), or peptidyl O-acyl hydroxamates (E Shaw, "CysteinyI Proteases and Their Selective Inactivation), pp 271-347). Other inhibitors are known in the art.

#### B. Screening Methods

25 Another aspect of the present invention is a method for screening a compound for anti-sortase-transamidase activity. This is an important aspect of the present invention, because it provides a method for screening for compounds that disrupt the sorting process and thus have potential antibiotic activity against Gram-positive bacteria.

30 In general, this method comprises the steps of: (1) providing an active fraction of sortase-transamidase enzyme; (2) performing an assay for sortase-transamidase activity in the presence and in the absence of the compound being screened; and (3) comparing the activity of the sortase-transamidase enzyme in the presence and in the absence of the compound.

35 The active fraction of sortase-transamidase enzyme can be a substantially purified sortase-transamidase enzyme preparation according to the

present invention, but can be a less purified preparation, such as a partially purified particulate preparation as described below.

5 The enzymatic activity can be measured by the cleavage of a suitable substrate, such as the construct having the Staphylococcal Enterotoxin B (SEB) gene fused to the cell wall sorting signal of Staphylococcal Protein A (SPA). The cleavage can be determined by monitoring the molecular weight of the products by sodium dodecyl sulfate-polyacrylamide gel electrophoresis or by other methods.

One particularly preferred assay for sortase-transamidase activity is the following:

10 Staphylococcal soluble RNA (sRNA) is prepared from *S. aureus* by a modification of the technique of Zubay (G. Zubay, J. Mol. Biol. 4: 347-356 (1962)). An overnight culture of *S. aureus* is diluted 1:10 in TSB and incubated at 37°C for 3 hr. The cells are harvested by centrifugation at 6000 rpm for 15 min.

For every gram of wet cell pellets, 2 ml of 0.01 M magnesium acetate,  
15 0.001 M Tris, pH 7.5 is used to suspend the pellets. The cell pellets are beaten by glass bead beater for 45 minutes in 5 minute intervals. The suspension is centrifuged twice at 2500 rpm for 5 minutes to remove the glass beads, then 0.5 ml phenol is added to the suspension. The suspension is vigorously shaken for 90 minutes at 4°C, and then centrifuged at 18,000 x g for 15 minutes. The nucleic acids in the top layer  
20 are precipitated by addition of 0.1 volume of 20% potassium acetate and 2 volumes of ethanol, then stored at 4°C for at least 36 hours. The precipitate is obtained by centrifugation at 5,000 x g for 5 minutes. Cold NaCl (1 ml) is added to the precipitate and stirred at 4°C for 1 hour. The suspension is centrifuged at 15,000 x g for 30 minutes. The sediments are washed with 0.5 ml of cold 1 M NaCl. The supernatants  
25 are combined and 2 volumes of ethanol is added to precipitate the tRNA. The precipitate is suspended in 0.1 ml of 0.2 M glycine, pH 10.3 and incubated for 3 hr at 37°C. This suspension is then made 0.4 M in NaCl and the RNA is precipitated by addition of 2 volumes of ethanol. The precipitate is dissolved in 0.7 ml of 0.3 M sodium acetate, pH 7.0. To this is slowly added 0.5 volume of isopropyl alcohol, with  
30 stirring. The precipitate is removed by centrifugation at 8,000 x g for 5 min. This precipitate is redissolved in 0.35 ml of 0.3 M sodium acetate, pH 7.0. To this is added 0.5 volume of isopropyl alcohol, using the same procedure as above. The precipitate is also removed by centrifugation. The combined supernatants from the two centrifugations are treated further with 0.37 ml of isopropyl alcohol. The resulting  
35 precipitate is dissolved in 75 µl of water and dialyzed against water overnight at 4°C. This sRNA is used in the sortase-transamidase assay.

Particulate sortase-transamidase enzyme is prepared for use in the assay by a modification of the procedure of Chatterjee & Park (A.N. Chatterjee & J.T. Park, Proc. Natl. Acad. Sci. USA 51: 9-16 (1964)). An overnight culture of *S. aureus* OS2 is diluted 1:50 in TSB and incubated at 37°C for 3 hr. Cells are harvested by centrifugation at 6000 rpm for 15 minutes, and washed twice with ice-cold water. The cells are disrupted by shaking 7 ml of 1 3% suspension of cells in 0.05 M Tris-HCl buffer, pH 7.5, 0.1 mM MgCl<sub>2</sub>, and 1 mM 2-mercaptoethanol with an equal volume of glass beads for 10-15 minutes in a beater. The glass beads are removed by centrifugation at 2000 rpm for 5 minutes. The crude extract is then centrifuged at 15,000 x g for 5 minutes. The supernatant is centrifuged again at 100,000 x g for 30 minutes. The light yellow translucent pellet is resuspended in 2 to 4 ml of 0.02 M Tris-HCl buffer, pH 7.5, containing 0.1 mM MgCl<sub>2</sub> and 1 mM 2-mercaptoethanol. This suspension represents the crude particulate enzyme and is used in the reaction mixture below.

The supernatant from centrifugation at 100,000 x g is passed through gel filtration using a Sephadex® G-25 agarose column (Pharmacia) to remove endogenous substrates. This supernatant is also used in the reaction mixture.

The complete reaction mixture contains in a final volume of 30 µl (M. Matsushashi et al., Proc. Natl. Acad. Sci. USA 54: 587-594 (1965)): 3 µmol of Tris-HCl, pH 7.8; 0.1 µmol of MgCl<sub>2</sub>; 1.3 µmol of KCl; 2.7 nmol of [<sup>3</sup>H] glycine (200 µCi/µmol); 2 nmol of UDP-M-pentapeptide; 5 nmol of UDP-N-acetylglucosamine; 0.2 µmol of ATP; 0.05 µmol of potassium phosphoenolpyruvate; 2.05 µg of chloramphenicol; 5 µg of pyruvate kinase; 0.025 µmol of 2-mercaptoethanol; 50 µg of staphylococcal sRNA prepared as above; 4 µg (as protein) of supernatant as prepared above; 271 µg of particulate enzyme prepared as above; and 8 nmol of a synthesized soluble peptide (HHHHHHAQALEPTGEENPF) (SEQ ID NO: 32) as a substrate.

The mixture is incubated at 20°C for 60 minutes. The mixture is then heated at 100°C for 1 minute. The mixture is diluted to 1 ml and precipitated with 50 µl nickel resin, and washed with wash buffer (1% Triton X-100, 0.1% sodium dodecyl sulfate, 50 mM Tris, pH 7.5). The nickel resin beads are counted in a scintillation counter to determine <sup>3</sup>H bound to the beads.

The effectiveness of the compound being screened to inhibit the activity of the sortase-transamidase enzyme can be determined by adding it to the assay mixture in a predetermined concentration and determining the resulting degree of inhibition of enzyme activity that results. Typically, a dose-response curve is generated using a range of concentrations of the compound being screened.



-23-

The particulate enzyme preparation of sortase-transamidase employed in this protocol can be replaced with any other sortase-transamidase preparation, purified or crude, staphylococcal, recombinant, or from any other source from any other Gram-positive bacterium as described above.

5           The soluble peptide is captured in this embodiment by its affinity for nickel resin as a result of the six histidine residues. More than six histidine residues can be used in the peptide. As an alternative, the soluble peptide can be captured by an affinity resulting from other interactions, such as streptavidin-biotin, glutathione S-transferase-glutathione, maltose binding protein-amylose, and the like, by  
10 replacing the six histidine residues with the amino acid sequence that constitutes the binding site in the peptide and employing the appropriate solid phase affinity resin containing the binding partner. Suitable peptides can be prepared by solid phase peptide synthesis using techniques well known in the art, such as those described in M. Bodanszky, "Peptide Chemistry: A Practical Textbook" (2d ed., Springer-Verlag,  
15 Berlin, 1993). For example, if the glutathione S-transferase-glutathione interaction is used, the active site of glutathione S-transferase (D.B. Smith & K.S. Johnson, "Single-Step Purification of Polypeptides Expressed in *Escherichia coli* as Fusions with Glutathione S-Transferase," *Gene* 67: 31-40 (1988)) can be substituted for the six histidine residues, and glutathione can be bound to the solid support.

20

#### IV. USE OF SORTASE-TRANSAMIDASE FOR PROTEIN AND PEPTIDE DISPLAY

##### A. Methods for Protein and Peptide Display

25           The sortase-transamidase enzyme of the present invention can also be used in a method of displaying a polypeptide on the surface of a gram-positive bacterium.

          In general, a first embodiment of this method comprises the steps of:  
(1) expressing a polypeptide having a sorting signal at its carboxyl-terminal end as described above; (2) forming a reaction mixture including: (i) the expressed  
30 polypeptide; (ii) a substantially purified sortase-transamidase enzyme; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide; and (3) allowing the sortase-transamidase to catalyze a reaction that cleaves the polypeptide within the LPX<sub>3</sub>X<sub>4</sub>G motif of the sorting signal and covalently cross-links the amino-terminal portion of the cleaved polypeptide to  
35 the peptidoglycan to display the polypeptide on the surface of the Gram-positive bacterium.

In this method, the polypeptide having the sorting signal at its carboxy-terminal end need not be expressed in a Gram-positive bacterium; it can be expressed in another bacterial system such as *Escherichia coli* or *Salmonella typhimurium*, or in a eukaryotic expression system.

5           The other method for protein targeting and display relies on direct expression of the chimeric protein in a Gram-positive bacterium and the action of the sortase-transamidase on the expressed protein. In general, such a method comprises the steps of: (1) cloning a nucleic acid segment encoding a chimeric protein into a Gram-positive bacterium to generate a cloned chimeric protein including therein a  
10       carboxyl-terminal sorting signal as described above, the chimeric protein including the polypeptide to be displayed; (2) growing the bacterium into which the nucleic acid segment has been cloned to express the cloned chimeric protein to generate a chimeric protein including therein a carboxyl-terminal sorting signal; and (3) covalent binding  
15       of the chimeric protein to the cell wall by the enzymatic action of the sortase-transamidase involving cleavage of the chimeric protein within the LPX<sub>3</sub>X<sub>4</sub>G motif so that the protein is displayed on the surface of the gram-positive bacterium in such a way that the protein is accessible to a ligand.

Typically, the Gram-positive bacterium is a species of *Staphylococcus*. A particularly preferred species of *Staphylococcus* is *Staphylococcus aureus*.

20           However, other Gram-positive bacteria such as *Streptococcus pyogenes*, other *Streptococcus* species, and Gram-positive bacteria of other genera can also be used.

Cloning the nucleic acid segment encoding the chimeric protein into the Gram-positive bacterium is performed by standard methods. In general, such  
25       cloning involves: (1) isolation of a nucleic acid segment encoding the protein to be sorted and covalently linked to the cell wall; (2) joining the nucleic acid segment to the sorting signal; (3) cloning by insertion into a vector compatible with the Gram-positive bacterium in which expression is to take place; and (4) incorporation of the vector including the new chimeric nucleic acid segment into the bacterium.

30           Typically, the nucleic acid segment encoding the protein to be sorted is DNA; however, the use of RNA in certain cloning steps is within the scope of the present invention.

When dealing with genes from eukaryotic organisms, it is preferred to use cDNA, because the natural gene typically contains intervening sequences or  
35       introns that are not translated. Alternatively, if the amino acid sequence is known, a synthetic gene encoding the protein to be sorted can be constructed by standard solid-phase oligodeoxyribonucleotide synthesis methods, such as the phosphotriester or

phosphite triester methods. The sequence of the synthetic gene is determined by the genetic code, by which each naturally occurring amino acid is specified by one or more codons. Additionally, if a portion of the protein sequence is known, but the gene or messenger RNA has not been isolated, the amino acid sequence can be used to construct a degenerate set of probes according to the known degeneracy of the genetic code. General aspects of cloning are described, for example, in J. Sambrook et al., "Molecular Cloning: A Laboratory Manual" (2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989); in B. Perbal, "A Practical Guide to Molecular Cloning" (2d ed., John Wiley & Sons, New York 1988), in S.L. Berger & A.R. Kimmel, "Guide to Molecular Cloning Techniques" (Methods in Enzymology, vol. 152, Academic Press, Inc., San Diego, 1987), and in D.V. Goeddel, ed., "Gene Expression Technology" (Methods in Enzymology, vol. 185, Academic Press, Inc., San Diego, 1991).

Once isolated, DNA encoding the protein to be sorted is then joined to the sorting signal. This is typically accomplished through ligation, such as using *Escherichia coli* or bacteriophage T4 ligase. Conditions for the use of these enzymes are well known and are described, for example, in the above general references.

The ligation is done in such a way so that the protein to be sorted and the sorting signal are joined in a single contiguous reading frame so that a single protein is produced. This may, in some cases, involve addition or deletion of bases of the cloned DNA segment to maintain a single reading frame. This can be done by using standard techniques.

Cloning is typically performed by inserting the cloned DNA into a vector containing control elements to allow expression of the cloned DNA. The vector is then incorporated into the bacterium in which expression is to occur, using standard techniques of transformation or other techniques for introducing nucleic acids into bacteria.

One suitable cloning system for *S. aureus* places the cloned gene under the control of the BlaZRI regulon (P.Z. Wang et al., Nucl. Acids Res. 19:4000 (1991)). Vectors and other cloning techniques for use in *Staphylococcus aureus* are described in B. Nilsson & L. Abrahmsen, "Fusion to Staphylococcal Protein A," in Gene Expression Technology, supra, p.144-161.

If the chimeric protein is cloned under control of the BlaZRI regulon, expression can be induced by the addition of the  $\beta$ -lactam antibiotic methicillin.

Another aspect of the present invention is a polypeptide displayed on the surface of a Gram-positive bacterium by covalent linkage of an amino-acid sequence of LPX<sub>3</sub>X<sub>4</sub> derived from cleavage of an LPX<sub>3</sub>X<sub>4</sub>G motif, as described above.

Yet another aspect of the present invention is a covalent complex comprising: (1) the displayed polypeptide; and (2) an antigen or hapten covalently cross-linked to the polypeptide.

5                   B. Screening Methods

These polypeptides associated with the cell surfaces of Gram-positive bacteria can be used in various ways for screening. For example, samples of expressed proteins from an expression library containing expressed proteins on the surfaces of the cells can be used to screen for clones that express a particular desired  
10 protein when a labeled antibody or other labeled specific binding partner for that protein is available.

These methods are based on the methods for protein targeting and display described above.

A first embodiment of such a method comprises: (1) expressing a  
15 cloned polypeptide as a chimeric protein having a sorting signal at its carboxy-terminal end as described above; (2) forming a reaction mixture including: (i) the expressed chimeric protein; (ii) a substantially purified sortase-transamidase enzyme; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide through the sorting signal; (3) binding of the  
20 chimeric protein covalently to the cell wall by the enzymatic action of a sortase-transamidase expressed by the Gram-positive bacterium involving cleavage of the chimeric protein within the LPX<sub>3</sub>X<sub>4</sub>G motif so that the polypeptide is displayed on the surface of the Gram-positive bacterium in such a way that the polypeptide is accessible to a ligand; and (4) reacting the displayed polypeptide with a labeled  
25 specific binding partner to screen the chimeric protein for reactivity with the labeled specific binding partner.

The nucleic acid segment encoding the chimeric protein is formed by methods well known in the art and can include a spacer.

In the last step, the cells are merely exposed to the labeled antibody or  
30 other labeled specific binding partner, unreacted antibodies removed as by a wash, and label associated with the cells detected by conventional techniques such as fluorescence, chemiluminescence, or autoradiography.

A second embodiment of this method employs expression in a Gram-positive bacterium that also produces a sortase-transamidase enzyme. This method  
35 comprises: (1) cloning a nucleic acid segment encoding a chimeric protein into a Gram-positive bacterium to generate a cloned chimeric protein including therein a carboxyl-terminal sorting signal as described above, the chimeric protein including

-27-

the polypeptide whose expression is to be screened; (2) growing the bacterium into which the nucleic acid segment has been cloned to express the cloned chimeric protein to generate a chimeric protein including therein a carboxyl-terminal sorting signal; (3) binding the polypeptide covalently to the cell wall by the enzymatic action of a  
5 sortase-transamidase expressed by the Gram-positive bacterium involving cleavage of the chimeric protein within the LPX<sub>3</sub>X<sub>4</sub>G motif so that the polypeptide is displayed on the surface of the Gram-positive bacterium in such a way that the polypeptide is accessible to a ligand; and (4) reacting the displayed polypeptide with a labeled specific binding partner to screen the chimeric protein for reactivity with the labeled  
10 specific binding partner.

#### V. USE OF SORTED MOLECULES FOR DIAGNOSIS AND TREATMENT OF BACTERIAL INFECTIONS

Sorted molecules can also be used for the diagnosis and treatment of  
15 bacterial infections caused by Gram-positive bacteria. Antibiotic molecules or fluorescent or any other diagnostic molecules can be chemically linked to a sorted peptide segment, which may include a spacer as described above, and then can be injected into animals or humans. These molecules are then sorted by the sortase-transamidase so that they are covalently linked to the cell wall of the bacteria.

20 In general, these methods comprise: (1) conjugating an antibiotic or a detection reagent to a protein including therein a carboxyl-terminal sorting signal to produce a conjugate; and (2) introducing the conjugate to an organism infected with a Gram-positive bacterium in order to cause the conjugate to be sorted and covalently cross-linked to the cell walls of the bacterium in order to treat or diagnose the  
25 infection.

The antibiotic used can be, but is not limited to, a penicillin, ampicillin, vancomycin, gentamicin, streptomycin, a cephalosporin, amikacin, kanamycin, neomycin, paromomycin, tobramycin, ciprofloxacin, clindamycin, rifampin, chloramphenicol, or norfloxacin, or a derivative of these antibiotics.

30 The detection reagent is typically an antibody or other specific binding partner labeled with a detectable label, such as a radiolabel. Such methods are well known in the art and need not be described further here.

Accordingly, another aspect of the present invention is a conjugate comprising an antibiotic or a detection reagent covalently conjugated to a protein  
35 including therein a carboxyl-terminal sorting signal as described above to produce a conjugate.

-28-

Yet another aspect of the present invention is a composition comprising the conjugate and a pharmaceutically acceptable carrier.

In this context, the conjugates can be administered using conventional modes of administration, including, but not limited to, intravenous, intraperitoneal, oral, or intralymphatic. Other routes of administration can alternatively be used. Oral or intraperitoneal administration is generally preferred. The composition can be administered in a variety of dosage forms, which include, but are not limited to, liquid solutions or suspensions, tablets, pills, powders, suppositories, polymeric microcapsules or microvesicles, liposomes, and injectable or infusible solutions. The preferred form depends on the mode of administration and the quantity administered.

The compositions for administration preferably also include conventional pharmaceutically acceptable carriers and adjuvants known in the art such as human serum albumin, ion exchangers, alumina, lecithin, buffered substances such as phosphate, glycine, sorbic acid, potassium sorbate, and salts or electrolytes such as protamine sulfate. The most effective mode of administration and dosage regimen for the conjugates as used in the methods in the present invention depend on the severity and course of the disease, the patient's health, the response to treatment, the particular strain of bacteria infecting the patient, other drugs being administered and the development of resistance to them, the accessibility of the site of infection to blood flow, pharmacokinetic considerations such as the condition of the patient's liver and/or kidneys that can affect the metabolism and/or excretion of the administered conjugates, and the judgment of the treating physician. According, the dosages should be titrated to the individual patient.

## 25 VI. USE OF SORTED POLYPEPTIDES FOR PRODUCTION OF VACCINES

Additionally, the sorted polypeptides covalently crosslinked to the cell walls of Gram-positive bacteria according to the present invention have a number of uses. One use is use in the production of vaccines that can be used to generate immunity against infectious diseases affecting mammals, including both human and non-human mammals, such as cattle, sheep, and goats, as well as other animals such as poultry and fish. This invention is of special importance to mammals. The usefulness of these complexes for vaccine production lies in the fact that the proteins are on the surface of the cell wall and are accessible to the medium surrounding the bacterial cells, so that the antigenic part of the chimeric protein is accessible to the antigen processing system. It is well known that presenting antigens in particulate form greatly enhances the immune response. In effect, bacteria containing antigenic peptides on the surfaces linked to the bacteria by these covalent interactions function

-29-

as natural adjuvants. Here follows a representative list of typical microorganisms that express polypeptide antigens against which useful antibodies can be prepared by the methods of the present invention:

(1) Fungi: *Candida albicans*, *Aspergillus fumigatus*, *Histoplasma capsulatum* (all cause disseminating disease), *Microsporium canis* (animal ringworm).

(2) Parasitic protozoa: (1) *Plasmodium falciparum* (malaria), *Trypanosoma cruzi* (sleeping sickness).

(3) Spirochetes: (1) *Borrelia bergdorferi* (Lyme disease), *Treponema pallidum* (syphilis), *Borrelia recurrentis* (relapsing fever), *Leptospira icterohaemorrhagiae* (leptospirosis).

(4) Bacteria: *Neisseria gonorrhoeae* (gonorrhea), *Staphylococcus aureus* (endocarditis), *Streptococcus pyogenes* (rheumatic fever), *Salmonella typhosa* (salmonellosis), *Hemophilus influenzae* (influenza), *Bordetella pertussis* (whooping cough), *Actinomyces israelii* (actinomycosis), *Streptococcus mutans* (dental caries), *Streptococcus equi* (strangles in horses), *Streptococcus agalactiae* (bovine mastitis), *Streptococcus anginosus* (canine genital infections).

(5) Viruses: Human immunodeficiency virus (HIV), poliovirus, influenza virus, rabies virus, herpes virus, foot and mouth disease virus, psittacosis virus, paramyxovirus, myxovirus, coronavirus.

Typically, the resulting immunological response occurs by both humoral and cell-mediated pathways. One possible immunological response is the production of antibodies, thereby providing protection against infection by the pathogen.

This method is not limited to protein antigens. As discussed below, non-protein antigens or haptens can be covalently linked to the C-terminal cell-wall targeting segment, which can be produced as an independently expressed polypeptide, either alone, or with a spacer at its amino-terminal end. If a spacer at the amino-terminal end is used, typically the spacer will have a conformation allowing the efficient interaction of the non-protein antigen or hapten with the immune system, most typically a random coil or  $\alpha$ -helical form. The spacer can be of any suitable length; typically, it is in the range of about 5 to about 30 amino acids; most typically, about 10 to about 20 amino acids. In this version of the embodiment, the independently expressed polypeptide, once expressed, can then be covalently linked to the hapten or non-protein antigen. Typical non-protein antigens or haptens include drugs, including both drugs of abuse and therapeutic drugs, alkaloids, steroids, carbohydrates, aromatic compounds, including many pollutants, and other compounds

-30-

that can be covalently linked to protein and against which an immune response can be raised.

Alternatively, a protein antigen can be covalently linked to the independently expressed cell-wall targeting segment or a cell-wall targeting segment including a spacer.

Many methods for covalent linkage of both protein and non-protein compounds to proteins are well known in the art and are described, for example, in P. Tijssen, "Practice and Theory of Enzyme Immunoassays" (Elsevier, Amsterdam, 1985), pp. 221-295, and in S.S. Wong, "Chemistry of Protein Conjugation and Cross-Linking" (CRC Press, Inc., Boca Raton, FL, 1993).

Many reactive groups on both protein and non-protein compounds are available for conjugation.

For example, organic moieties containing carboxyl groups or that can be carboxylated can be conjugated to proteins via the mixed anhydride method, the carbodiimide method, using dicyclohexylcarbodiimide, and the N-hydroxysuccinimide ester method.

If the organic moiety contains amino groups or reducible nitro groups or can be substituted with such groups, conjugation can be achieved by one of several techniques. Aromatic amines can be converted to diazonium salts by the slow addition of nitrous acid and then reacted with proteins at a pH of about 9. If the organic moiety contains aliphatic amines, such groups can be conjugated to proteins by various methods, including carbodiimide, tolylene-2,4-diisocyanate, or maleimide compounds, particularly the N-hydroxysuccinimide esters of maleimide derivatives. An example of such a compound is 4-(N-maleimidomethyl)-cyclohexane-1-carboxylic acid. Another example is m-maleimidobenzoyl-N-hydroxysuccinimide ester. Still another reagent that can be used is N-succinimidyl-3-(2-pyridyldithio) propionate. Also, bifunctional esters, such as dimethylpimelimidate, dimethyladipimidate, or dimethylsuberimidate, can be used to couple amino-group-containing moieties to proteins.

Additionally, aliphatic amines can also be converted to aromatic amines by reaction with p-nitrobenzoylchloride and subsequent reduction to a p-aminobenzoylamide, which can then be coupled to proteins after diazotization.

Organic moieties containing hydroxyl groups can be cross-linked by a number of indirect procedures. For example, the conversion of an alcohol moiety to the half ester of succinic acid (hemisuccinate) introduces a carboxyl group available for conjugation. The bifunctional reagent sebacoyldichloride converts alcohol to acid chloride which, at pH 8.5, reacts readily with proteins. Hydroxyl-containing organic



-31-

moieties can also be conjugated through the highly reactive chlorocarbonates, prepared with an equal molar amount of phosgene.

For organic moieties containing ketones or aldehydes, such carbonyl-containing groups can be derivatized into carboxyl groups through the formation of O-(carboxymethyl) oximes. Ketone groups can also be derivatized with p-hydrazinobenzoic acid to produce carboxyl groups that can be conjugated to the specific binding partner as described above. Organic moieties containing aldehyde groups can be directly conjugated through the formation of Schiff bases which are then stabilized by a reduction with sodium borohydride.

One particularly useful cross-linking agent for hydroxyl-containing organic moieties is a photosensitive noncleavable heterobifunctional cross-linking reagent, sulfosuccinimidyl 6-[4'-azido-2'-nitrophenylamino] hexanoate. Other similar reagents are described in S.S. Wong, "Chemistry of Protein Conjugation and Cross-Linking," *supra*.

Other cross-linking reagents can be used that introduce spacers between the organic moiety and the specific binding partner.

These methods need not be described further here.

## VII. PRODUCTION OF SUBSTANTIALLY PURIFIED SORTASE-TRANSAMIDASE ENZYME

Another aspect of the present invention is methods for the production of substantially purified sortase-transamidase enzyme.

### A. Methods Involving Expression of Cloned Gene

One method for the production of substantially purified sortase-transamidase enzyme involves the expression of the cloned gene. The isolation of the nucleic acid segment or segments encoding the sortase-transamidase enzyme is described above; these nucleic acid segment or segments are then incorporated into a vector and then use to transform a host in which the enzyme can be expressed. In one alternative, the host is a Gram-positive bacterium.

The next step in this alternative is expression in a Gram-positive bacterium to generate the cloned sortase-transamidase enzyme. Expression is typically under the control of various control elements associated with the vector incorporating the DNA encoding the sortase-transamidase gene; such elements can include promoters and operators, which can be regulated by proteins such as repressors. The conditions required for expression of cloned proteins in gram-positive bacteria, particularly *S. aureus*, are well known in the art and need not be

-32-

further recited here. An example is the induction of expression of lysostaphin under control of the BlaZRI regulon induced by the addition of methicillin.

When expressed in *Staphylococcus aureus*, the chimeric protein is typically first exported with an amino-terminal leader peptide, such as the hydrophobic signal peptide at the amino-terminal region of the cloned lysostaphin of Recsei et al. (P. Recsei et al., "Cloning, Sequence, and Expression of the Lysostaphin Gene from *Staphylococcus simulans*," Proc. Natl. Acad. Sci. USA 84:1127-1131 (1987)).

Alternatively, the cloned nucleic acid segment encoding the sortase-transamidase enzyme can be inserted in a vector that contains sequences allowing expression of the sortase-transamidase in another organism, such as *E. coli* or *S. typhimurium*. A suitable host organism can then be transformed or transfected with the vector containing the cloned nucleic acid segment. Expression is then performed in that host organism.

The expressed enzyme is then purified using standard techniques. Techniques for the purification of cloned proteins are well known in the art and need not be detailed further here. One particularly suitable method of purification is affinity chromatography employing an immobilized antibody to sortase. Other protein purification methods include chromatography on ion-exchange resins, gel electrophoresis, isoelectric focusing, and gel filtration, among others.

One particularly useful form of affinity chromatography for purification of cloned proteins, such as sortase-transamidase, as well as other proteins, such as glutathione S-transferase and thioredoxin, that have been extended with carboxyl-terminal histidine residues, is chromatography on a nickel-sepharose column. This allows the purification of a sortase-transamidase enzyme extended at its carboxyl terminus with a sufficient number of histidine residues to allow specific binding of the protein molecule to the nickel-sepharose column through the histidine residues. The bound protein is then eluted with imidazole. Typically, six or more histidine residues are added; preferably, six histidine residues are added. One way of adding the histidine residues to a cloned protein, such the sortase-transamidase, is through PCR with a primer that includes nucleotides encoding the histidine residues. The histidine codons are CAU and CAC expressed as RNA, which are CAT and CAC as DNA. Amplification of the cloned DNA with appropriate primers will add the histidine residues to yield a new nucleic acid segment, which can be recloned into an appropriate host for expression of the enzyme extended with the histidine residues.

### B. Other Methods

Alternatively, the sortase-transamidase can be purified from Gram-positive bacteria by standard methods, including precipitation with reagents such as ammonium sulfate or protamine sulfate, ion-exchange chromatography, gel filtration chromatography, affinity chromatography, isoelectric focusing, and gel electrophoresis, as well as other methods known in the art.

Because the sortase-transamidase is a cysteine protease, one particularly useful method of purification involves covalent chromatography by thiol-disulfide interchange, using a two-protonic-state gel containing a 2-mercaptopyridine leaving group, such as Sepharose 2B-glutathione 2-pyridyl disulfide or Sepharose 6B-hydroxypropyl 2-pyridyl disulfide. Such covalent chromatographic techniques are described in K. Brocklehurst et al., "Cysteine Proteases," in New Comprehensive Biochemistry, Volume 16: Hydrolytic Enzymes (A. Neuberger & K. Brocklehurst, eds., Elsevier, New York, 1987), ch. 2, pp. 39-158.

## VIII. FURTHER APPLICATIONS OF SORTASE-TRANSAMIDASE

### A. Production of Antibodies

Antibodies can be prepared to the substantially purified sortase-transamidase of the present invention, whether the sortase-transamidase is purified from bacteria or produced from recombinant bacteria as a result of gene cloning procedures. Because the substantially purified enzyme according to the present invention is a protein, it is an effective antigen, and antibodies can be made by well-understood methods such as those disclosed in E. Harlow & D. Lane, "Antibodies: A Laboratory Manual" (Cold Spring Harbor Laboratory, 1988). In general, antibody preparation involves immunizing an antibody-producing animal with the protein, with or without an adjuvant such as Freund's complete or incomplete adjuvant, and purification of the antibody produced. The resulting polyclonal antibody can be purified by techniques such as affinity chromatography.

Once the polyclonal antibodies are prepared, monoclonal antibodies can be prepared by standard procedures, such as those described in Chapter 6 of Harlow & Lane, supra.

### B. Derivatives for Affinity Chromatography

Another aspect of the present invention is derivatives of the cloned, substantially purified sortase-transamidase of the present invention extended at its carboxyl terminus with a sufficient number of histidine residues to allow specific binding of the protein molecule to a nickel-sepharose column through the histidine

-34-

residues. Typically, six or more histidine residues are added; preferably, six histidine residues are added.

The histidine residues can be added to the carboxyl terminus through PCR cloning as described above.

5 This invention is further described by means of the following example. This Example is for illustrative purposes only, and are not to be construed as limiting the scope of the invention in any manner.

### EXAMPLE

#### 10 Identification of Sortase-Transamidase Generation of ts Mutants through Chemical Mutagenesis

To create random mutations in the chromosome, *Staphylococcus aureus* strain OS2 (RN4220 erm spa-) was mutagenized by exposure to the DNA-modifying agent N-methyl-N-nitro-N-nitrosoguanidine. Cultures were incubated  
15 with the mutagen for varying periods of time, then placed on TSB agar plates to measure viability. Cultures were subsequently plated on TSB+ rifampicin (10 µg/ml) to determine the mutation frequency based on resistance to the single target site antibiotic. Once a maximum mutation frequency was reached, cell cultures were exposed to two successive rounds of a penicillin selection (5 µg/ml at 42°C) to enrich  
20 for mutants that had a growth effect by lysis of the cells growing at this temperature. Mutants were screened for growth at 30°C and 42°C by streaking individual colonies on TSB agar plates at the permissive and non-permissive temperatures, respectively. These colonies that demonstrated a growth defect at the non-permissive temperature were rechecked at 42°C, and subsequently frozen at -80°C in a 5% BSA, 5%  
25 monosodium glutamate (MSG) solution. In this manner, a collection of temperature-sensitive mutants was assembled.

#### Transformation and Screening of ts Mutants

In order to isolate mutants that demonstrated a defect in the surface  
30 display of protein, it was necessary to develop a screening process to locate these strains. Previous studies that had indicated a typical secretory-dependent process, in conjunction with known C-terminal cleavage of translocated proteins, were used to elucidate a selection scheme to isolate the desirable mutants. The construct harboring the staphylococcal enterotoxin B (SEB) gene fused to the cell wall sorting signal of  
35 staphylococcal protein A (SPA) was used in this assay. This reporter molecule has been shown to be properly processed not only by the secretory machinery and through signal peptidase cleavage of an N-terminal secretion signal, but also to be correctly

-35-

sorted after secretion to its peptidoglycan substrate. Mutant cells were made competent using 0.5 M sucrose washes, and were transformed with the staphylococcal/*Escherichia coli* shuttle vector pOS 1 containing the SEB-SPA reporter through electroporation. Transformants were selected by virtue of their resistance to chloramphenicol as encoded by the plasmid. The cells were then screened for properties indicative of a defect in the processing of precursor molecules to a fully matured and anchored surface protein.

Cultured cells were induced at 42°C, and then pulsed with S-35 ProMix (80% Met, 20% Cys), in order to label all synthesized proteins. Samples were precipitated through acid treatment, then digested with lysostaphin (100 µg/ml) and subsequently reprecipitated. This was followed by solubilization in hot sodium dodecyl sulfate (4%), and an immunoprecipitation with anti-SEB antibodies. Samples were run on sodium dodecyl sulfate-polyacrylamide gel electrophoresis and finally exposed to phosphorimager evaluation. The resulting banding patterns were analyzed and quantitated.

This procedure yielded three anti-SEB reacting species, termed P1, P2 and M. P1, the largest precursor which migrated to 33 kDa, represents the complete gene product encoded by the SEB-SPA construct, with no modifications. P2, the second precursor, most likely represents the product after cleavage of the amino-terminal signal peptide found in SEB, and thus migrates to 32 kDa. The smallest species, M, is a lysostaphin-solubilized, maturely anchored peptide that has neither the signal peptide nor the remainder of the carboxyl-terminal sorting signal after the cleavage. This band migrates at approximately 29 kDa.

Analysis of these species was conducted through phosphorimager quantitation, and mutants were selected based upon the proposed phenotype that the inhibition in sorting would result in an accumulation of P2 and a reduction in the production of M. Through this process, two mutants, SM317 and SM329, were earmarked for further analysis due to their elevated ratio of P2/M over wild-type.

Both mutants finally demonstrated an accumulation of P2 after a 5 min kinetic analysis, but also clearly showed a decrease in the mutant's ability to degrade the species over time as measured by samples chased with cold methionine. These results were interpreted to mean that the ability of the mutants to process mature cell wall anchored peptides was impaired, quite possibly due to the less efficient activity of the sortase-transamidase enzyme.

Protoplasts were made of these mutants in order to cure them of their plasmids encoding the reporter construct, and subsequently retransformed with the SEB/SPA containing plasmid to once again test for preservation of this phenotype.

-36-

Upon a favorable result, the mutants were prepared for complementation by a *S. aureus* chromosomal library.

#### Generation of an *S. aureus* Chromosomal Library and Complementation of ts Defect

5                   A staphylococcal library was made through a Sau3AI digest of the chromosomal DNA preparation from *S. aureus* strain RN4220. DNA was isolated through a phenol-chloroform extraction from lysed cells, and digested for various times until the correct partial digest pattern was observed. Fragments greater than 2.5 kb were inserted into the BamHI cloning site in the multi-cloning sequence (MCS) of plasmid pC194-MCS. This heterogeneous mixture of plasmids was then transformed  
10 into competent OS2 cells. Approximately 15,000 clones were harvested. DNA was prepared and transformed into competent cells made of both mutants, and simultaneously plated at 30°C and 42°C to screen for complementation of the ts mutant phenotype.

15                   Through this process, four chromosomal inserts from each mutant were found to complement the ts phenotype by conferring growth at 42°C. Due to the nature of the mutagenesis, it is at this point necessary to demonstrate definitively that the ts defect is somehow linked to the defect in processing. This is done through illustration that the plasmids harboring the chromosomal inserts not only complement  
20 the temperature sensitivity, but also relieve the accumulation of P2 at the expense of M in these mutants. Therefore, the complemented mutants were screened along with the non-complemented versions against wild type OS2.

#### Screen for Sorting Defect Complementation

25                   This assay was conducted differently for each of the two mutants. SM-317 was screened by the insertion of the SEB/SPA fragment aboard the replication defective pCL 84 vector that possessed integration capability into the chromosome of *S. aureus* cells. The site-specific integration, mediated by the integrase gene supplied in trans by pCL112, disrupts the lipase gene, which can be  
30 assayed for by the lack of hydrolysis on egg yolk agar plates. Once successfully integrated, the RN4220 chromosomal fragments that complement the ts mutation can be added to make the cells ready for screening.

                    SM329 was assayed by another approach. The pC194 plasmid harboring the complementing stretch of DNA was fused to an *E. coli* replicon  
35 pHSG399 that contained the SEB/SPA gene. The shuttle vector thus provided both a reporter substrate as well as a complementing activity.

The result of this screen demonstrated that over several time periods of the pulse-chase, the complementing insert added to the mutant reduced the accretion of P2 from the elevated level characteristic of the mutant to the wild type range in both mutants, the more dramatically in SM-329. This may be due to the fact that the substrate for sorting in SM-317 is found in only one copy per cell, whereas in SM-329, the reporter aboard pC194 is present at approximately 15 copies. Nevertheless, these results indicate that these mutants are in the sortase-transamidase gene and the sequencing of both chromosomal inserts was therefore undertaken. It should be pointed out that the complementing activity of each of the respective mutants was not transferable to the other, neither in terms of temperature sensitivity nor for the processing defect. Also, the four complementing clones isolated from each mutant seemed to behave in an identical manner, and also seemed to possess very similar restriction sites when digested with various specific endonucleases. Therefore, one clone was chosen from each mutant for sequencing.

#### Sequencing and Characterization of the *S. aureus* Complementing Determinants

The chromosomal inserts carrying the sorting defect complementing capabilities were sequenced. This work was done completely by automated sequence analysis using dideoxyribonucleotides. Sequence data was confirmed by duplicate analysis of both strands of DNA. Comparison was done to all known nucleotide and protein sequence was currently found in the GenBank service. The partial, crude sequence of the *S. aureus* gene is shown in Figure 10 (SEQ ID NO: 28 & 29). The partial carboxy-terminal amino acid sequence of the open reading frame generated from the gene sequence of Figure 10 (SEQ ID NO: 28 & 29) is shown in Figure 11 (SEQ ID NO: 2).

Several stretches of high homology were found, to both known and putative proteins of varying function. The first 364 bases of the SM-317 complementing gene insert been identified as encoding a protein that is a homologue of a putative *Bacillus* peptidase in the GCVT-SPOIIIAA intergenic region (GenBank Accession No. 1731048; Y. Kobayashi et al.). The sequence of this putative peptidase is shown in Figure 5 (SEQ ID NO: 3) and its hydrophobicity profile is shown in Figure 6. The hydrophobicity is calculated according to the method of J. Kyte & R.F. Doolittle, "A Simple Method for Displaying the Hydropathic Character of a Protein," J. Mol. Biol. 157: 105-132 (1982). To a lesser degree of homology, the protein encoded by this complementing gene insert is homologous to aminopeptidase P of *Lactococcus lactis* (GenBank Accession No. 1915907; J. Matos). The amino acid sequence of this aminopeptidase is shown in Figure 7 (SEQ ID NO: 4). To a still

lesser degree of homology, the protein encoded by this complementing gene insert is homologous to the proline dipeptidase of *Lactobacillus delbrueckii lactis* (GenBank Accession No. 1172066; K. Stucky et al., "Cloning and DNA Sequence Analysis of pepQ, a Prolidase Gene from *Lactobacillus delbrueckii* subsp. *lactis* and Partial Characterization of Its Product," Mol. Gen. Genet. 247: 494-500 (1995). The amino acid sequence of this proline dipeptidase is shown in Figure 8 (SEQ ID NO: 5).

An additional complementing gene insert, in a vector designated pCOMP1, has also been sequenced. The DNA sequence of this complementing gene insert is shown in Figure 12 (SEQ ID NO: 30), together with the amino acid sequence (SEQ ID NO: 31) of the protein translated from this DNA sequence, and the hydrophobicity profile of the protein translated from the DNA sequence is shown in Figure 13. The amino acid sequence of the protein translated from the sequence of Figure 12 (SEQ ID NOS: 30 & 31) has virtually no homology with the amino acid sequence of the protein shown in Figure 11 (SEQ ID NO: 2). In particular, the amino acid sequence of the protein of Figure 12 (SEQ ID NOS: 30 & 31) has a single cysteine residue.

Although Applicants do not intend to be bound by this theory, the existence of these two complementing inserts that define different polypeptide sequences suggests that the sortase-transamidase enzyme of *S. aureus* is a heterooligomer, with two or more different subunits that have different amino acid sequences. Mutations in at least two different subunits can give rise to the temperature sensitive phenotype and can then be complemented for. Alternatively, the synthesis of the peptidoglycan may require additional enzymes.

Upon completion of the sequencing, and a study of all open reading frames, candidate genes were selected for further analysis. These genes were expressed in the mutants to determine if they complement both the ts and sorting defects. Upon success in this capacity, the genes were disrupted in wild-type *S. aureus* to determine their essentiality and possible biological roles.

### 30 Materials and Methods

Mutagenesis of *S. aureus* Strain OS2. *S. aureus* strain OS2 (RN4220 erm spa-) was mutagenized by treatment with N-methyl-N-nitro-N-nitrosoguanidine at 2 mg/ml. Culture OD was measured at 660 nm for viability. After 45 min, cultures were spun down at 4,000 rpm for 10 min, and washed with citrate buffer, pH 5.5. Cells were resuspended in citrate buffer to a concentration of 5x10<sup>8</sup> cells/ml. These cultures were serially diluted in TSB (tryptic soy broth) and



-39-

plated on TSB+ rifampicin (10 µg/ml) agar plates. Mutation frequencies were determined to be  $5 \times 10^5$  mutation/cfu.

Enrichment and Selection of ts Mutants. A 1:100 dilution of cell cultures grown overnight at 30°C was added to TSB, and allowed to grow for 2 hr, at which time penicillin G was added at 5 µg/ml. Culture viability was measured by taking OD readings at 660 nm at various times until the concentration dropped to a stable point. Cells were washed twice in TSB, and the enrichment was repeated a second time.

After two successive rounds of penicillin selection for growth arrest at 42°C, individual colonies were picked and simultaneously streaked on duplicate TSB plates incubated at 30°C and 42°C. Those colonies that had a growth defect at 42°C were rechecked at this nonpermissive temperature, and stored at -80°C in a 5% BSA 5% MSG solution.

Transformation of Competent Cells. Mutant cells were made competent by diluting overnight cultures 1:10 in TSB, and growing to 0.3 OD at 660 nm. Cells were then spun down at 7,500 rpm for 15 min and resuspended in an equal volume of 0.5 M sucrose. After another pelleting, cells were resuspended in 0.5 volume sucrose, and incubated for 30 min at 4°C. Following another spin, cells were brought up in 0.1 volume sucrose. These competent cells were then transformed with the appropriate plasmids encoding chloramphenicol (10 µg/ml) resistance by electroporation at 200 ohms, 25 µF, and 2.5 kV in 0.2 cm cuvettes. Cells were plated on TSB plus chloramphenicol incubated at 30°C. Transformed strains were frozen in BSA/MSG at -80°C.

Pulse-Chase Screen of Mutants. Strains were inoculated in chemically defined media with chloramphenicol and grown overnight at 30°C. Cultures were diluted 1:10 into medium IV (O. Schneewind et al., "Sorting of Protein A to the Staphylococcal Cell Wall," *Cell* 70: 267-281 (1992)), grown for 3 hr, and then induced at 42°C for 20 min. At this time, cultures were pulse labeled with 50 µCi of S-35 ProMix for 5 min, and then terminated with 5% trichloroacetic acid (TCA). Cells were incubated at 4° for 30 min, centrifuged at 12,500 rpm for 15 min, and the supernatants aspirated. After resuspension in acetone, cells were spun again and aspirated to dryness. At this time, cells were treated with lysostaphin (100 µg/ml) for 30 min or until noticeable clearing, and subjected to another TCA/acetone precipitation. After lysis of cells by boiling for 10 min with 4% SDS in 0.5M Tris, pH 8.0, proteins were immunoprecipitated with anti-SEB for 1 hr and protein A-Sepharose beads for another 1 hr. Samples were washed three times in RIPA buffer, pH 8.0 (300 mM NaCl, 2% Triton X-100, 1% deoxycholate, 0.2% SDS), and protein

-40-

was solubilized with SDS-urea sample buffer (50 mM Tris-HCl, pH 6.8, 3% SDS, 5% 2-mercaptoethanol, 3.5 M urea), with boiling for 10 min. Samples were run on SDS-polyacrylamide gel electrophoresis gels and exposed to a phosphorimager screen overnight. Quantitations were done on ImageQuant software.

- 5                    DNA Sequencing. DNA was sequenced on a Perkin-Elmer automated sequencer after PCR using dye-terminating ready reaction mixed with SS-Taq polymerase. GenBank analysis was done using BLAST software to search the database.

10                    ADVANTAGES OF THE PRESENT INVENTION

- In isolating and characterizing the gene for the *S. aureus* sortase-transamidase enzyme, we have determined the existence of a new site for antibiotic action that can be used to screen new antibiotics active against Gram-positive pathogens, such as *Staphylococcus*, *Actinomyces*, *Mycobacterium*, *Streptococcus*,  
15    *Bacillus*, and other medically important Gram-positive pathogens increasingly resistant to conventional antibiotics. The availability of substantially purified *S. aureus* sortase-transamidase enzyme provides a method of screening compounds for inhibition of the enzyme.

- The purified sortase-transamidase enzyme of the present invention  
20    also yields a method of surface display of peptides and proteins that has advantages over phage display, as well as providing methods for producing vaccines against a large variety of antigens that can be covalently bound to the surfaces of Gram-positive bacteria.

- Although the present invention has been described with considerable  
25    detail, with reference to certain preferred versions thereof, other versions and embodiments are possible. Therefore, the scope of the invention is determined by the following claims.

We claim:

1. A substantially purified sortase-transamidase enzyme from a Gram-positive bacterium, the enzyme catalyzing a reaction that covalently cross-links the carboxyl terminus of a protein having a sorting signal to the peptidoglycan of a Gram-positive bacterium, the sorting signal having a motif of LPX<sub>3</sub>X<sub>4</sub>G therein, wherein sorting occurs by cleavage between the fourth and fifth residues of the LPX<sub>3</sub>X<sub>4</sub>G motif.
2. The substantially purified sortase-transamidase enzyme of claim 1 wherein the Gram-positive bacterium is a species selected from the group consisting of *Staphylococcus aureus*, *S. sobrinus*, *Enterococcus faecalis*, *Streptococcus pyogenes*, and *Listeria monocytogenes*.
3. The substantially purified sortase-transamidase enzyme of claim 2 wherein the Gram-positive bacterium is *Staphylococcus aureus*.
4. The substantially purified sortase-transamidase enzyme of claim 1 wherein one subunit of the enzyme has a molecular weight of about 41,000 daltons.
5. The substantially purified sortase-transamidase enzyme of claim 4 wherein the sorting signal further comprises: (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine.
6. The enzyme of claim 1 wherein the enzyme includes therein an amino acid sequence selected from the group consisting of: (1) D-P-K-L-K-E-I-Y-Q-I-V-L-E-S-Q-M-K-A-I-N-E-I-R-P-G-M-T-G-A-E-A-D-A-I-S-R-N-Y-L-E-S-K-G-Y-G-K-E-F-G-H-S-L-G-H-G-I-G-L-E-I-H-E-G-P-M-L-A-R-T-I-Q-D-K-L-Q-V-N-N-C-V-T-V-E-P-G-V-Y-I-E-G-L-G-I-R-I-E-D-D-I-L-I-T-E-N-G-C-Q-V-F-T-K-C-T-K-D-L-I-V-L-T (SEQ ID NO: 2); (2) M-V-K-V-T-D-Y-S-N-S-K-L-G-K-E-I-A-P-E-V-L-

-42-

S-V-I-A-S-I-A-T-S-E-V-E-G-I-T-G-H-F-A-E-L-K-E-T-N-L-E-K-V-S-R-K-N-L-S-R-D-L-K-I-E-S-K-E-G-I-Y-I-D-V-Y-C-A-L-K-H-G-V-N-I-S-K-T-A-N-K-I-Q-T-S-I-F-N-S-I-S-N-M-T-A-I-E-P-K-Q-I-N-I-H-I-T-Q-I-V-I-E-K (SEQ ID NO: 31); and (3) sequences incorporating one or more conservative amino acid substitutions in SEQ ID NO:2 or SEQ ID NO: 31, wherein the conservative amino acid substitutions are any of the following: (1) any of isoleucine, leucine, and valine for any other of these amino acids; (2) aspartic acid for glutamic acid and vice versa; (3) glutamine for asparagine and vice versa; and (4) serine for threonine and vice versa.

10

7. The enzyme of claim 6 wherein the amino acid sequence is D-P-K-L-K-E-I-Y-Q-I-V-L-E-S-Q-M-K-A-I-N-E-I-R-P-G-M-T-G-A-E-A-D-A-I-S-R-N-Y-L-E-S-K-G-Y-G-K-E-F-G-H-S-L-G-H-G-I-G-L-E-I-H-E-G-P-M-L-A-R-T-I-Q-D-K-L-Q-V-N-N-C-V-T-V-E-P-G-V-Y-I-E-G-L-G-I-R-I-E-D-D-I-L-I-T-E-N-G-C-Q-V-F-T-K-C-T-K-D-L-I-V-L-T (SEQ ID NO: 2).

15

8. The enzyme of claim 6 wherein the amino acid sequence is M-V-K-V-T-D-Y-S-N-S-K-L-G-K-E-I-A-P-E-V-L-S-V-I-A-S-I-A-T-S-E-V-E-G-I-T-G-H-F-A-E-L-K-E-T-N-L-E-K-V-S-R-K-N-L-S-R-D-L-K-I-E-S-K-E-G-I-Y-I-D-V-Y-C-A-L-K-H-G-V-N-I-S-K-T-A-N-K-I-Q-T-S-I-F-N-S-I-S-N-M-T-A-I-E-P-K-Q-I-N-I-H-I-T-Q-I-V-I-E-K (SEQ ID NO: 31).

20

25 9. A nucleic acid sequence encoding the enzyme of claim 6.

10. A nucleic acid sequence encoding the enzyme of claim 7.

11. A nucleic acid sequence encoding the enzyme of claim 8.

30

12. A nucleic acid sequence encoding a substantially purified sortase-transamidase enzyme from a Gram-positive bacterium, the enzyme having a subunit with a molecular weight of about 41,000 daltons and catalyzing a reaction that covalently cross-links the carboxyl terminus of a protein having a sorting signal to the peptidoglycan of a Gram-positive bacterium, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged

35

-43-

residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein  $X_3$  is any of the twenty naturally-occurring L-amino acids and  $X_4$  is selected from the group consisting of alanine, serine, and threonine, and wherein sorting occurs by cleavage between the fourth and fifth residues of the  $LPX_3X_4G$  motif, wherein the nucleic acid sequence includes therein a sequence selected from the group consisting of: (1)

GATCCTAAACTGAAAGAAATATATCAAATAGTACTTGAATCTCAAATGAA  
 AGCAATTAATGAGATTAGACCTGGCATGACTGGTGCAGAAGCTGATGCCA  
 10 TTTCAAGAACTATTTAGAGTCAAAAGGGTATGGAAAAGAATTTGGACAT  
 TCACTAGGACATGGTATTGGTTTAGAAATCCATGAAGGGCCAATGCTGGC  
 TCGTACGATACAAGATAAACTTCAAGTTAACAACGTGTGTTACAGTAGAAC  
 CTGGTGTATTATAGAAAGGTTTGGGCGGTATAAGAATAGAAGATGATATT  
 TTAATTACAGAAAATGGTTGTCAAGTCTTTACTAAATGCACAAAAGACCTT  
 15 ATAGTTTAAACATAA (SEQ ID NO: 28); (2)  
 ATGGTCAAAGTAACTGATTATTCAAATTCAAATTAGGTAAAGTAGAAAT  
 AGCGCCAGAAGTGCTATCTGTTATTGCAAGTATAGCTACTTCGGAAGTCG  
 AAGGCATCACTGGCCATTTTGCTGAATTAAGAAACAAATTTAGAAAAA  
 GTTAGTCGTAAAAATTTAAGCCGTGATTTAAAAATCGAGAGTAAAGAAGA  
 20 TGGCATATATATAGATGTATATTGTGCATTAAACATGGTAATATTTCAAA  
 AACTGCAAACAAAATTCAAACGTCAATTTTAAATTCAATTTCTAATATGAC  
 AGCGATAGAACCTAAGCAAATTAATATTCACATTACACAAATCGTTATTG  
 AAAAGTAA (SEQ ID NO: 30); or (3) a sequence complementary to SEQ ID NO: 28  
 or SEQ ID NO: 30.

25

13. A nucleic acid sequence encoding a substantially purified sortase-transamidase enzyme from a Gram-positive bacterium, the enzyme having a subunit with a molecular weight of about 41,000 daltons and catalyzing a reaction that covalently cross-links the carboxyl terminus of a protein having a sorting signal to the peptidoglycan of a Gram-positive bacterium, the sorting signal having (1) a motif of  $LPX_3X_4G$  therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein  $X_3$  is any of the twenty naturally-occurring L-amino acids and  $X_4$  is selected from the group consisting of alanine, serine, and threonine, and wherein sorting occurs by cleavage between the fourth and

35

-44-

fifth residues of the LPX<sub>3</sub>X<sub>4</sub>G motif, wherein the nucleic acid sequence hybridizes with a sequence selected from the group consisting of: (1)

GATCCTAAACTGAAAGAAATATATCAAATAGTACTTGAATCTCAAATGAA  
AGCAATTAATGAGATTAGACCTGGCATGACTGGTGCAGAAGCTGATGCCA  
5 TTTCAAGAAACTATTTAGAGTCAAAAGGGTATGGAAAAGAATTTGGACAT  
TCACTAGGACATGGTATTGGTTTAGAAATCCATGAAGGGCCAATGCTGGC  
TCGTACGATACAAGATAAACTTCAAGTTAACAACGTGTGTTACAGTAGAAC  
CTGGTGTTTATATAGAAGGTTTGGGCGGTATAAGAATAGAAGATGATATT  
TTAATTACAGAAAATGGTTGTCAAGTCTTTACTAAATGCACAAAAGACCTT  
10 ATAGTTTAAACATAA (SEQ ID NO: 28); (2)  
ATGGTCAAAGTAACTGATTATTCAAATTCAAAATTAGGTAAAGTAGAAAT  
AGCGCCAGAAGTGCTATCTGTTATTGCAAGTATAGCTACTTCGGAAGTCG  
AAGGCATCACTGGCCATTTTGCTGAATTAAGAAACAAATTTAGAAAAA  
GTTAGTCGTAAAAATTTAAGCCGTGATTTAAAAATCGAGAGTAAAGAAGA  
15 TGGCATATATATAGATGTATATTGTGCATTAAACATGGTAATATTTCAA  
AACTGCAACAAAATTCAAACGTCAATTTTAAATTCAATTTCTAATATGAC  
AGCGATAGAACCTAAGCAAATTAATATTCACATTACACAAATCGTTATTG  
AAAAGTAA (SEQ ID NO: 30) or (3) a sequence complementary to SEQ ID NO: 28  
or SEQ ID NO: 30, with no greater than about a 15% mismatch under stringent  
20 conditions.

14. The nucleic acid sequence of claim 13 wherein the mismatch is no greater than about 5%.

25 15. The nucleic acid sequence of claim 14 wherein the mismatch is no greater than about 2%.

16. A vector comprising the nucleic acid sequence of claim 9 operatively linked to at least one control sequence that controls the expression or  
30 regulation of the nucleic acid sequence.

17. A vector comprising the nucleic acid sequence of claim 10 operatively linked to at least one control sequence that controls the expression or regulation of the nucleic acid sequence.

35

-45-

18. A vector comprising the nucleic acid sequence of claim 11 operatively linked to at least one control sequence that controls the expression or regulation of the nucleic acid sequence.

5           19. A vector comprising the nucleic acid sequence of claim 12 operatively linked to at least one control sequence that controls the expression or regulation of the nucleic acid sequence.

10           20. A vector comprising the nucleic acid sequence of claim 13 operatively linked to at least one control sequence that controls the expression or regulation of the nucleic acid sequence.

21. A host cell transfected with the vector of claim 16.

15           22. A host cell transfected with the vector of claim 17.

23. A host cell transfected with the vector of claim 18.

20           24. A host cell transfected with the vector of claim 19.

25. A host cell transfected with the vector of claim 20.

26. A method for producing a substantially purified sortase-transamidase enzyme comprising the steps of:

25           (a) culturing the host cell of claim 21 under conditions in which the host cell expresses the encoded sortase-transamidase enzyme; and  
             (b) purifying the expressed enzyme to produce substantially purified sortase-transamidase enzyme.

30           27. A method for producing a substantially purified sortase-transamidase enzyme comprising the steps of:

             (a) culturing the host cell of claim 22 under conditions in which the host cell expresses the encoded sortase-transamidase enzyme; and  
             (b) purifying the expressed enzyme to produce substantially purified  
35 sortase-transamidase enzyme.

-46-

28. A method for producing a substantially purified sortase-transamidase enzyme comprising the steps of:

(a) culturing the host cell of claim 23 under conditions in which the host cell expresses the encoded sortase-transamidase enzyme; and

5 (b) purifying the expressed enzyme to produce substantially purified sortase-transamidase enzyme.

29. A method for producing a substantially purified sortase-transamidase enzyme comprising the steps of:

10 (a) culturing the host cell of claim 24 under conditions in which the host cell expresses the encoded sortase-transamidase enzyme; and

(b) purifying the expressed enzyme to produce substantially purified sortase-transamidase enzyme.

15 30. A method for producing a substantially purified sortase-transamidase enzyme comprising the steps of:

(a) culturing the host cell of claim 25 under conditions in which the host cell expresses the encoded sortase-transamidase enzyme; and

20 (b) purifying the expressed enzyme to produce substantially purified sortase-transamidase enzyme.

31. Substantially purified sortase-transamidase enzyme produced by the process of claim 26.

25 32. Substantially purified sortase-transamidase enzyme produced by the process of claim 27.

33. Substantially purified sortase-transamidase enzyme produced by the process of claim 28.

30

34. Substantially purified sortase-transamidase enzyme produced by the process of claim 29.

35 35. Substantially purified sortase-transamidase enzyme produced by the process of claim 30.



-47-

36. A method for screening a compound for anti-sortase-transamidase activity comprising the steps of:

(a) providing the substantially purified sortase-transamidase enzyme of claim 1;

5 (b) performing an assay for sortase-transamidase in the presence and in the absence of the compound; and

(c) comparing the activity of the sortase-transamidase enzyme in the presence and in the absence of the compound to screen the compound for sortase-transamidase activity.

10

37. A method for screening a compound for anti-sortase-transamidase activity comprising the steps of:

(a) providing the substantially purified sortase-transamidase enzyme of claim 3;

15 (b) performing an assay for sortase-transamidase in the presence and in the absence of the compound; and

(c) comparing the activity of the sortase-transamidase enzyme in the presence and in the absence of the compound to screen the compound for sortase-transamidase activity.

20

38. A method for screening a compound for anti-sortase-transamidase activity comprising the steps of:

(a) providing the substantially purified sortase-transamidase enzyme of claim 31;

25 (b) performing an assay for sortase-transamidase in the presence and in the absence of the compound; and

(c) comparing the activity of the sortase-transamidase enzyme in the presence and in the absence of the compound to screen the compound for sortase-transamidase activity.

30

39. A method for screening a compound for anti-sortase-transamidase activity comprising the steps of:

(a) providing the substantially purified sortase-transamidase enzyme of claim 32;

35 (b) performing an assay for sortase-transamidase in the presence and in the absence of the compound; and

-48-

(c) comparing the activity of the sortase-transamidase enzyme in the presence and in the absence of the compound to screen the compound for sortase-transamidase activity.

5                   40. A method for screening a compound for anti-sortase-transamidase activity comprising the steps of:

(a) providing the substantially purified sortase-transamidase enzyme of claim 33;

10                   (b) performing an assay for sortase-transamidase in the presence and in the absence of the compound; and

(c) comparing the activity of the sortase-transamidase enzyme in the presence and in the absence of the compound to screen the compound for sortase-transamidase activity.

15                   41. A method for screening a compound for anti-sortase-transamidase activity comprising the steps of:

(a) providing the substantially purified sortase-transamidase enzyme of claim 34;

20                   (b) performing an assay for sortase-transamidase in the presence and in the absence of the compound; and

(c) comparing the activity of the sortase-transamidase enzyme in the presence and in the absence of the compound to screen the compound for sortase-transamidase activity.

25                   42. A method for screening a compound for anti-sortase-transamidase activity comprising the steps of:

(a) providing the substantially purified sortase-transamidase enzyme of claim 35;

30                   (b) performing an assay for sortase-transamidase in the presence and in the absence of the compound; and

(c) comparing the activity of the sortase-transamidase enzyme in the presence and in the absence of the compound to screen the compound for sortase-transamidase activity.

-49-

43. A method for screening a compound for anti-sortase-transamidase activity comprising the steps of:

(a) providing an active fraction of sortase-transamidase enzyme from a Gram-positive bacterium;

5 (b) performing an assay for sortase-transamidase in the presence and in the absence of the compound; and

(c) comparing the activity of the sortase-transamidase enzyme in the presence and in the absence of the compound to screen the compound for sortase-transamidase activity.

10

44. The method of claim 43 wherein the active fraction of sortase-transamidase enzyme is a particulate fraction from *Staphylococcus aureus*.

45. The method of claim 43 wherein the assay for sortase-transamidase enzyme is performed by monitoring the capture of a soluble peptide that is a substrate for the enzyme by its interaction with an affinity resin.

15

46. The method of claim 45 wherein the soluble peptide includes a sequence of at least six histidine residues and the affinity resin contains nickel.

20

47. The method of claim 45 wherein the soluble peptide includes the active site of glutathione S-transferase and the affinity resin contains glutathione.

48. The method of claim 45 wherein the soluble peptide includes the active site of streptavidin and the affinity resin contains biotin.

25

49. The method of claim 45 wherein the soluble peptide includes the active site of maltose binding protein and the affinity resin contains amylose.

50. An antibody specifically binding the substantially purified sortase-transamidase enzyme of claim 1.

30

51. An antibody specifically binding the substantially purified sortase-transamidase enzyme of claim 3.

35

52. An antibody specifically binding the substantially purified sortase-transamidase enzyme of claim 31.

53. An antibody specifically binding the substantially purified sortase-transamidase enzyme of claim 32.

5 54. An antibody specifically binding the substantially purified sortase-transamidase enzyme of claim 33.

55. An antibody specifically binding the substantially purified sortase-transamidase enzyme of claim 34.

10 56. An antibody specifically binding the substantially purified sortase-transamidase enzyme of claim 35.

15 57. A protein molecule comprising the substantially purified sortase-transamidase enzyme of claim 1 extended at its carboxyl-terminus with a sufficient number of histidine residues to allow specific binding of the protein molecule to a nickel-sepharose column through the histidine residues added at the carboxyl-terminus.

20 58. A protein molecule comprising the substantially purified sortase-transamidase enzyme of claim 3 extended at its carboxyl-terminus with a sufficient number of histidine residues to allow specific binding of the protein molecule to a nickel-sepharose column through the histidine residues added at the carboxyl-terminus.

25 59. A protein molecule comprising the substantially purified sortase-transamidase enzyme of claim 31 extended at its carboxyl-terminus with a sufficient number of histidine residues to allow specific binding of the protein molecule to a nickel-sepharose column.

30 60. A protein molecule comprising the substantially purified sortase-transamidase enzyme of claim 32 extended at its carboxyl-terminus with a sufficient number of histidine residues to allow specific binding of the protein molecule to a nickel-sepharose column.

35

61. A protein molecule comprising the substantially purified sortase-transamidase enzyme of claim 33 extended at its carboxyl-terminus with a sufficient number of histidine residues to allow specific binding of the protein molecule to a nickel-sepharose column.

5

62. A protein molecule comprising the substantially purified sortase-transamidase enzyme of claim 34 extended at its carboxyl-terminus with a sufficient number of histidine residues to allow specific binding of the protein molecule to a nickel-sepharose column.

10

63. A protein molecule comprising the substantially purified sortase-transamidase enzyme of claim 35 extended at its carboxyl-terminus with a sufficient number of histidine residues to allow specific binding of the protein molecule to a nickel-sepharose column.

15

64. A method for displaying a polypeptide on the surface of a Gram-positive bacterium comprising the steps of:

(a) expressing a polypeptide having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;

25

(b) forming a reaction mixture including: (i) the expressed polypeptide; (ii) the substantially purified sortase-transamidase of claim 1; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide; and

30

(c) allowing the sortase-transamidase to catalyze a reaction that cleaves the polypeptide within the LPX<sub>3</sub>X<sub>4</sub> motif of the sorting signal and covalently cross-links the amino-terminal portion of the cleaved polypeptide to the peptidoglycan to display the polypeptide on the surface of the Gram-positive bacterium.

35

-52-

65. A method for displaying a polypeptide on the surface of a Gram-positive bacterium comprising the steps of:

- (a) expressing a polypeptide having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;
- (b) forming a reaction mixture including: (i) the expressed polypeptide; (ii) the substantially purified sortase-transamidase of claim 3; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide; and
- (c) allowing the sortase-transamidase to catalyze a reaction that cleaves the polypeptide within the LPX<sub>3</sub>X<sub>4</sub>G motif of the sorting signal and covalently cross-links the amino-terminal portion of the cleaved polypeptide to the peptidoglycan to display the polypeptide on the surface of the Gram-positive bacterium.

66. A method for displaying a polypeptide on the surface of a Gram-positive bacterium comprising the steps of:

- (a) expressing a polypeptide having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;
- (b) forming a reaction mixture including: (i) the expressed polypeptide; (ii) the substantially purified sortase-transamidase enzyme of claim 31; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide; and
- (c) allowing the sortase-transamidase to catalyze a reaction that cleaves the polypeptide within the LPX<sub>3</sub>X<sub>4</sub>G motif of the sorting signal and covalently cross-links the amino-terminal portion of the cleaved polypeptide to the

peptidoglycan to display the polypeptide on the surface of the Gram-positive bacterium.

67. A method for displaying a polypeptide on the surface of a Gram-positive bacterium comprising the steps of:

- (a) expressing a polypeptide having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;
- (b) forming a reaction mixture including: (i) the expressed polypeptide; (ii) the substantially purified sortase-transamidase enzyme of claim 32; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide; and
- (c) allowing the sortase-transamidase to catalyze a reaction that cleaves the polypeptide within the LPX<sub>3</sub>X<sub>4</sub>G motif of the sorting signal and covalently cross-links the amino-terminal portion of the cleaved polypeptide to the peptidoglycan to display the polypeptide on the surface of the Gram-positive bacterium.

68. A method for displaying a polypeptide on the surface of a Gram-positive bacterium comprising the steps of:

- (a) expressing a polypeptide having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;
- (b) forming a reaction mixture including: (i) the expressed polypeptide; (ii) the substantially purified sortase-transamidase enzyme of claim 33; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide; and

-54-

(c) allowing the sortase-transamidase to catalyze a reaction that cleaves the polypeptide within the LPX<sub>3</sub>X<sub>4</sub>G motif of the sorting signal and covalently cross-links the amino-terminal portion of the cleaved polypeptide to the peptidoglycan to display the polypeptide on the surface of the Gram-positive bacterium.

69. A method for displaying a polypeptide on the surface of a Gram-positive bacterium comprising the steps of:

(a) expressing a polypeptide having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;

(b) forming a reaction mixture including: (i) the expressed polypeptide; (ii) the substantially purified sortase-transamidase enzyme of claim 34; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide; and

(c) allowing the sortase-transamidase to catalyze a reaction that cleaves the polypeptide within the LPX<sub>3</sub>X<sub>4</sub>G motif of the sorting signal and covalently cross-links the amino-terminal portion of the cleaved polypeptide to the peptidoglycan to display the polypeptide on the surface of the Gram-positive bacterium.

70. A method for displaying a polypeptide on the surface of a Gram-positive bacterium comprising the steps of:

(a) expressing a polypeptide having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;



-55-

(b) forming a reaction mixture including: (i) the expressed polypeptide; (ii) the substantially purified sortase-transamidase enzyme of claim 35; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide; and

5 (c) allowing the sortase-transamidase to catalyze a reaction that cleaves the polypeptide within the LPX<sub>3</sub>X<sub>4</sub>G motif of the sorting signal and covalently cross-links the amino-terminal portion of the cleaved polypeptide to the peptidoglycan to display the polypeptide on the surface of the Gram-positive bacterium.

10

71. A method for displaying a polypeptide on the surface of a Gram-positive bacterium comprising the steps of:

(a) cloning a nucleic acid segment encoding a chimeric protein into a Gram-positive bacterium to generate a cloned chimeric protein including therein a  
15 carboxyl-terminal sorting signal, the chimeric protein including the polypeptide to be displayed, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues  
20 being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;

(b) growing the bacterium into which the nucleic acid segment has been cloned to express the cloned chimeric protein to generate a chimeric protein  
25 including therein a carboxyl-terminal sorting signal; and

(c) binding the polypeptide covalently to the cell wall by the enzymatic action of a sortase-transamidase expressed by the Gram-positive bacterium involving cleavage of the chimeric protein within the LPX<sub>3</sub>X<sub>4</sub>G motif so that the polypeptide is displayed on the surface of the Gram-positive bacterium in such a way that the  
30 polypeptide is accessible to a ligand.

72. A polypeptide displayed on the surface of a Gram-positive bacterium by covalent linkage of an amino-acid sequence of LPX<sub>3</sub>X<sub>4</sub> derived from cleavage of an LPX<sub>3</sub>X<sub>4</sub>G motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring  
35 L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine, the polypeptide being displayed on the surface of the Gram-positive bacterium in such a way that the polypeptide is accessible to a ligand.

73. A covalent complex comprising:

- (a) the polypeptide of claim 72; and
- (b) an antigen or hapten covalently cross-linked to the polypeptide.

5

74. The covalent complex of claim 73 wherein an antigen is covalently cross-linked to the polypeptide.

75. The covalent complex of claim 73 wherein a hapten is covalently cross-linked to the peptide.

10

76. A method for vaccination of an animal comprising the step of immunizing the animal with the displayed polypeptide of claim 72 to generate an immune response against the displayed polypeptide.

15

77. A method for vaccination of an animal comprising the step of immunizing the animal with the covalent complex of claim 73 to generate an immune response against the antigen or hapten of the covalent complex.

20

78. A method for screening for expression of a cloned polypeptide comprising the steps of:

- (a) expressing a cloned polypeptide as a chimeric protein having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;

30

- (b) forming a reaction mixture including: (i) the expressed chimeric protein; the substantially purified sortase-transamidase enzyme of claim 1; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide through the sorting signal;

35

- (c) binding the chimeric protein covalently to the cell wall by the enzymatic action of a sortase-transamidase expressed by the Gram-positive bacterium involving cleavage of the chimeric protein within the LPX<sub>3</sub>X<sub>4</sub>G motif so

-57-

that the polypeptide is displayed on the surface of the Gram-positive bacterium in such a way that the polypeptide is accessible to a ligand; and

(d) reacting the displayed polypeptide with a labeled specific binding partner to screen the chimeric protein for reactivity with the labeled specific binding partner.

79. A method for screening for expression of a cloned polypeptide comprising the steps of:

(a) expressing a cloned polypeptide as a chimeric protein having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;

(b) forming a reaction mixture including: (i) the expressed chimeric protein; (ii) the substantially purified sortase-transamidase enzyme of claim 3; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide through the sorting signal;

(c) binding the chimeric protein covalently to the cell wall by the enzymatic action of a sortase-transamidase expressed by the Gram-positive bacterium involving cleavage of the chimeric protein within the LPX<sub>3</sub>X<sub>4</sub>G motif so that the polypeptide is displayed on the surface of the Gram-positive bacterium in such a way that the polypeptide is accessible to a ligand; and

(d) reacting the displayed polypeptide with a labeled specific binding partner to screen the chimeric protein for reactivity with the labeled specific binding partner.

80. A method for screening for expression of a cloned polypeptide comprising the steps of:

(a) expressing a cloned polypeptide as a chimeric protein having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two

-58-

positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein  $X_3$  is any of the twenty naturally-occurring L-amino acids and  $X_4$  is selected from the group consisting of alanine, serine, and threonine;

- 5 (b) forming a reaction mixture including: (i) the expressed chimeric protein; (ii) the substantially purified sortase-transamidase enzyme of claim 31; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide through the sorting signal;

- 10 (c) binding the chimeric protein covalently to the cell wall by the enzymatic action of a sortase-transamidase expressed by the Gram-positive bacterium involving cleavage of the chimeric protein within the  $LPX_3X_4G$  motif so that the polypeptide is displayed on the surface of the Gram-positive bacterium in such a way that the polypeptide is accessible to a ligand; and

- 15 (d) reacting the displayed polypeptide with a labeled specific binding partner to screen the chimeric protein for reactivity with the labeled specific binding partner.

81. A method for screening for expression of a cloned polypeptide comprising the steps of:

- 20 (a) expressing a cloned polypeptide as a chimeric protein having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of  $LPX_3X_4G$  therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two  
25 positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein  $X_3$  is any of the twenty naturally-occurring L-amino acids and  $X_4$  is selected from the group consisting of alanine, serine, and threonine;

- 30 (b) forming a reaction mixture including: (i) the expressed chimeric protein; (ii) the substantially purified sortase-transamidase enzyme of claim 32; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide through the sorting signal;

- 35 (c) binding the chimeric protein covalently to the cell wall by the enzymatic action of a sortase-transamidase expressed by the Gram-positive bacterium involving cleavage of the chimeric protein within the  $LPX_3X_4G$  motif so that the polypeptide is displayed on the surface of the Gram-positive bacterium in such a way that the polypeptide is accessible to a ligand; and

(d) reacting the displayed polypeptide with a labeled specific binding partner to screen the chimeric protein for reactivity with the labeled specific binding partner.

5                   82. A method for screening for expression of a cloned polypeptide comprising the steps of:

                  (a) expressing a cloned polypeptide as a chimeric protein having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids  
10   carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-  
15   occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;

                  (b) forming a reaction mixture including: (i) the expressed chimeric protein; (ii) the substantially purified sortase-transamidase enzyme of claim 33; and (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide through the sorting signal;

20                   (c) binding the chimeric protein covalently to the cell wall by the enzymatic action of a sortase-transamidase expressed by the Gram-positive bacterium involving cleavage of the chimeric protein within the LPX<sub>3</sub>X<sub>4</sub>G motif so that the polypeptide is displayed on the surface of the Gram-positive bacterium in such a way that the polypeptide is accessible to a ligand; and

25                   (d) reacting the displayed polypeptide with a labeled specific binding partner to screen the chimeric protein for reactivity with the labeled specific binding partner.

30                   83. A method for screening for expression of a cloned polypeptide comprising the steps of:

                  (a) expressing a cloned polypeptide as a chimeric protein having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids  
35   carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-

-60-

occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;

- (b) forming a reaction mixture including: (i) the expressed chimeric protein; (ii) the substantially purified sortase-transamidase enzyme of claim 34; and  
5 (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide through the sorting signal;

- (c) binding the chimeric protein covalently to the cell wall by the enzymatic action of a sortase-transamidase expressed by the Gram-positive bacterium involving cleavage of the chimeric protein within the LPX<sub>3</sub>X<sub>4</sub>G motif so  
10 that the polypeptide is displayed on the surface of the Gram-positive bacterium in such a way that the polypeptide is accessible to a ligand; and

- (d) reacting the displayed polypeptide with a labeled specific binding partner to screen the chimeric protein for reactivity with the labeled specific binding partner.

15

84. A method for screening for expression of a cloned polypeptide comprising the steps of:

- (a) expressing a cloned polypeptide as a chimeric protein having a sorting signal at its carboxy-terminal end, the sorting signal having: (1) a motif of  
20 LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-  
25 occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;

- (b) forming a reaction mixture including: (i) the expressed chimeric protein; (ii) the substantially purified sortase-transamidase enzyme of claim 35; and  
30 (iii) a Gram-positive bacterium having a peptidoglycan to which the sortase-transamidase can link the polypeptide through the sorting signal;

- (c) binding the chimeric protein covalently to the cell wall by the enzymatic action of a sortase-transamidase expressed by the Gram-positive bacterium involving cleavage of the chimeric protein within the LPX<sub>3</sub>X<sub>4</sub>G motif so  
35 that the polypeptide is displayed on the surface of the Gram-positive bacterium in such a way that the polypeptide is accessible to a ligand; and

-61-

(d) reacting the displayed polypeptide with a labeled specific binding partner to screen the chimeric protein for reactivity with the labeled specific binding partner.

5                   85. A method for screening for expression of a cloned polypeptide comprising the steps of:

(a) cloning a nucleic acid segment encoding a chimeric protein into a Gram-positive bacterium to generate a cloned chimeric protein including therein a carboxyl-terminal sorting signal, the chimeric protein including the polypeptide  
10 whose expression is to be screened, the sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at  
15 residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine;

(b) growing the bacterium into which the nucleic acid segment has been cloned to express the cloned chimeric protein to generate a chimeric protein  
20 including therein a carboxyl-terminal sorting signal;

(c) binding the polypeptide covalently to the cell wall by the enzymatic action of a sortase-transamidase expressed by the Gram-positive bacterium involving cleavage of the chimeric protein within the LPX<sub>3</sub>X<sub>4</sub>G motif so that the polypeptide is displayed on the surface of the Gram-positive bacterium in such a way that the  
25 polypeptide is accessible to a ligand; and

(d) reacting the displayed polypeptide with a labeled specific binding partner to screen the chimeric protein for reactivity with the labeled specific binding partner.

30                   86. A method for the diagnosis or treatment of a bacterial infection caused by a Gram-positive bacterium comprising the steps of:

(a) conjugating an antibiotic or a detection reagent to a protein including therein a carboxyl-terminal sorting signal to produce a conjugate, the carboxyl-terminal sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a  
35 substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues

-62-

being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine; and

- (b) introducing the conjugate to an organism infected with a Gram-  
5 positive bacterium in order to cause the conjugate to be sorted and covalently cross-linked to the cell walls of the bacterium in order to treat or diagnose the infection.

87. The method of claim 86 wherein an antibiotic is conjugated to the  
protein.

10

88. The method of claim 87 wherein the antibiotic is selected from the group consisting of a penicillin, ampicillin, vancomycin, gentamicin, streptomycin, a cephalosporin, amikacin, kanamycin, neomycin, paromomycin, tobramycin, ciprofloxacin, clindamycin, rifampin, chloramphenicol, norfloxacin, and a derivative  
15 of these antibiotics.

89. The method of claim 86 wherein a detection reagent is conjugated to the protein.

20

90. A conjugate comprising an antibiotic or a detection reagent covalently conjugated to a protein including therein a carboxyl-terminal sorting signal to produce a conjugate, the carboxyl-terminal sorting signal having: (1) a motif of LPX<sub>3</sub>X<sub>4</sub>G therein; (2) a substantially hydrophobic domain of at least 31 amino acids carboxyl to the motif; and (3) a charged tail region with at least two positively charged  
25 residues carboxyl to the substantially hydrophobic domain, at least one of the two positively charged residues being arginine, the two positively charged residues being located at residues 31-33 from the motif, wherein X<sub>3</sub> is any of the twenty naturally-occurring L-amino acids and X<sub>4</sub> is selected from the group consisting of alanine, serine, and threonine.

30

91. The conjugate of claim 90 wherein an antibiotic is conjugated to the protein.

92. The conjugate of claim 91 wherein the antibiotic is selected from  
35 the group consisting of a penicillin, ampicillin, vancomycin, gentamicin, streptomycin, a cephalosporin, amikacin, kanamycin, neomycin, paromomycin,



-63-

tobramycin, ciprofloxacin, clindamycin, rifampin, chloramphenicol, norfloxacin, and a derivative of these antibiotics.

93. The conjugate of claim 90 wherein a detection reagent is  
5 conjugated to the protein.

94. A composition comprising:  
(a) the conjugate of claim 90; and  
(b) a pharmaceutically acceptable carrier.

10

95. A substantially purified protein having at least about 50% match  
with best alignment, in at least one subunit of the protein, with the amino acid  
sequences of at least one of the putative *Bacillus* peptidase (SEQ ID NO: 3), the  
aminopeptidase P of *Lactococcus lactis* (SEQ ID NO: 4), or the proline dipeptidase of  
15 *Lactobacillus delbrueckii lactis* (SEQ ID NO: 5) and having sortase-transamidase  
activity.

96. The substantially purified protein of claim 95 wherein the match  
with best alignment with the amino acid sequences of at least one of the putative  
20 *Bacillus* peptidase (SEQ ID NO: 3), the aminopeptidase P of *Lactococcus lactis* (SEQ  
ID NO: 4), or the proline dipeptidase of *Lactobacillus delbrueckii lactis* (SEQ ID NO:  
5) is at least about 60%.

97. The substantially purified protein of claim 96 wherein the match  
25 with best alignment with the amino acid sequences of at least one of the putative  
*Bacillus* peptidase (SEQ ID NO: 3), the aminopeptidase P of *Lactococcus lactis* (SEQ  
ID NO: 4), or the proline dipeptidase of *Lactobacillus delbrueckii lactis* (SEQ ID NO:  
5) is at least about 70%.

98. A substantially purified protein having sortase-transamidase  
30 activity and a hydrophobicity profile of at least one subunit of the protein, that,  
determined as the mean absolute value of the hydrophobicity difference per residue,  
differs from the hydrophobicity profile of a putative *Bacillus* peptidase (SEQ ID NO:  
3) by no more than about 2 units on the hydrophobicity scale.

35

-64-

99. The substantially purified protein of claim 98 wherein the hydrophobicity profile differs from the hydrophobicity profile of the putative *Bacillus* peptidase (SEQ ID NO: 3) by no more than about 1 unit.

5           100. The substantially purified protein of claim 99 wherein the hydrophobicity profile differs from the hydrophobicity profile of the putative *Bacillus* peptidase (SEQ ID NO: 3) by no more than about 0.5 unit.

10           101. A nucleic acid sequence encoding the substantially purified protein of claim 95.

            102. A nucleic acid sequence encoding the substantially purified protein of claim 98.

15           103. A vector comprising the nucleic acid sequence of claim 101 operatively linked to at least one control sequence that controls the expression or regulation of the nucleic acid sequence.

20           104. A vector comprising the nucleic acid sequence of claim 102 operatively linked to at least one control sequence that controls the expression or regulation of the nucleic acid sequence.

            105. A host cell transfected with the vector of claim 103.

25           106. A host cell transfected with the vector of claim 104.

            107. A method for producing a substantially purified protein having sortase-transamidase activity comprising the steps of:

- 30           (a) culturing the host cell of claim 105 under conditions in which the host cell expresses the protein having sortase-transamidase activity; and  
            (b) purifying the expressed protein to produce substantially purified protein having sortase-transamidase activity.

35           108. A method for producing a substantially purified protein having sortase-transamidase activity comprising the steps of:

- (a) culturing the host cell of claim 106 under conditions in which the host cell expresses the protein having sortase-transamidase activity; and

-65-

(b) purifying the expressed protein to produce substantially purified protein having sortase-transamidase activity.

5 109. A substantially purified protein having sortase-transamidase activity and a hydrophobicity profile of at least one subunit of the protein, that, determined as the mean absolute value of the hydrophobicity difference per residue, differs from the hydrophobicity profile of the sequence of SEQ ID NO: 31 by no more than about 2 units on the hydrophobicity scale.

10 110. The substantially purified protein of claim 109 wherein the hydrophobicity profile differs from the hydrophobicity profile of the sequence of SEQ ID NO: 31 by no more than about 1 unit.

15 111. The substantially purified protein of claim 110 wherein the hydrophobicity profile differs from the hydrophobicity profile of the sequence of SEQ ID NO: 31 by no more than about 0.5 unit.

20 112. A nucleic acid sequence encoding the substantially purified protein of claim 109.

113. A vector comprising the nucleic acid sequence of claim 112 operatively linked to at least one control sequence that controls the expression or regulation of the nucleic acid sequence.

25 114. A host cell transfected with the vector of claim 113.

115. A method for producing a substantially purified protein having sortase-transamidase activity comprising the steps of:

30 (a) culturing the host cell of claim 114 under conditions in which the host cell expresses the protein having sortase-transamidase activity; and

(b) purifying the expressed protein to produce substantially purified protein having sortase-transamidase activity.

1/18

FIG. 1

Matching Percentage (Total Window: 21%, Alignment Window: 22%)

S.aureus -6 .....DPK-----L-----K-EI-----YQ--- 43

S.pyogenes 1 MLQYSQKLPKKEFAMSGFLEQRLGHCLRQMAEKGLEALLVTHLTNSYILTG 50

44 -----I-----VLE--SOM-----K-AINE---I---R-P----- 93

51 FSGTAATVLITAKRRVLITDSRYTLLAKASVEGFDIIESRTPLKVVAELL 100

94 -----G-----MT-----G-----A-----E-----A--- 143

101 EADQIDCLGFEDQVSFSFYQAMQAELSGITLLAQSGFVEHLRLIKDASEI 150

144 D--A---IS-----R-----NYLE-----SK--- 193

151 DTIAKACISDKAFEDALDFIKPGTTTERDLANFLDFRMRQYGASGTSFD 200

194 -----GY-----G-----K-----E-----FG-----H-SLG-----H-G- 243

201 IIVASGYLSAMPHGRASDKVIONKESLTMDFGCYNNHYVSDMTRTIHIGQ 250

244 -----IG-LE-----IHE-----G-----PM-L-----A-----R-T 293

251 VTDEEREIYALVLAANKALIAKASAGMTYSDFDGI PRQLITEAGYGSRET 300

294 --IQ-----D-----K---L-Q---VNNCVTVEPGVYIEGLG---GI 343

301 HGIGHGIGLDIHENPFFGKSEQLLOAGMV---VTDEPGIY---LDNKYGV 350

344 RIEDDIL-ITENGCCQVFTKCTKDILVLT\*..... 393

351 RIEDD-LVITKTGCQVLTLPKELIVL\*..... 400

(SEQ ID No: 2 &amp; 34)

**FIG. 2**

[illegible]

(SEQ ID No: 2 & 34)

3/18

FIG. 3A

10	20	30	40	50	60
ATGCTACAAT	ATTCTCAAA	GTTACCAAAG	GAGTTCGCGA	TGTCAGGATT	TTTAGAACAA
70	80	90	100	110	120
CGATTAGGTC	ACTGCCTAAG	GCAGATGGCA	GAGAAGGGGC	TAGAGGCTCT	TCTAGTCACC
130	140	150	160	170	180
CATTTAACCA	ATAGTTATTA	CTTGACAGGT	TTTTCTGGAA	CTGCAGCAAC	TGTTTTGATA
190	200	210	220	230	240
ACGGCCAAAC	GTCGTGTTT	GATCACAGAT	TCACGTTATA	CCTTGCTTGC	TAAAGCTAGT
250	260	270	280	290	300
GTTGAGGGAT	TTGATATTAT	CGAAAGCCGC	ACGCCGCTTA	AGGTTGTGGC	AGAATTGTTA
310	320	330	340	350	360
GAGGCTGATC	AAATAGATTG	CCTTGGTTT	GAGGACCAGG	TATCGTTTTC	TTTTTACCAG
370	380	390	400	410	420
GCCATGCAAG	CAGAACTGTC	AGGAATAACC	TTGCTTGCTC	AGTCAGGTTT	TGTGGAGCAT
430	440	450	460	470	480
TTACGTCCTA	TTAAGGACGC	CTCTGAAATC	GATACCAATG	CTAAAGCGTG	CTCGATCTCA
490	500	510	520	530	540
GACAAAGCAT	TTGAAGATGC	TCTTGATTTT	ATTAAACCAG	GGACAACCCAC	TGAACGTGAC
550	560	570	580	590	600
CTGGCTAATT	TTTTAGATT	TCGTATGCCG	CAGTATGGTG	CCAGCGGCAC	ATCATTTGAT

4/18

FIG. 3B

610	620	630	640	650	660
ATCATTGTAG	CTTCAGGCTA	TCTCTCTGCC	ATGCCTCATG	GACGGGCCAG	TGACAAGGTT
670	680	690	700	710	720
ATCCAGAATA	AAGAGAGCTT	GACCATGGAC	TTTGGGTGTT	ACTACAATCA	CTATGTTAGT
730	740	750	760	770	780
GATATGACGA	GGACCAATTCA	TATTGGCCAA	GTTACTGATG	AAGAACGTGA	GATTTATGCT
790	800	810	820	830	840
CTTGTTCTTG	CTGCTAATAA	GGCTTTAATT	GCTAAAGCTA	GGCTGGCAT	GACTTATAGT
850	860	870	880	890	900
GACTTTGACG	GTATTCCGCG	CCAACTCATC	ACTGAGGCGG	GTTATGGCAG	TCGCTTCACA
910	920	930	940	950	960
CATGGCATTG	GTCATGGCAT	CGGGCTTGAC	ATCCATGAGA	ATCCATTTTT	TGGGAATCT
970	980	990	1000	1010	1020
GAGCAACTTC	TCCAAGCTGG	AATGGTGGTA	ACAGATGAGC	CAGGTATCTA	TTTGGATAAC
1030	1040	1050	1060	1070	1080
AAATATGGTG	TCCGTATTGA	AGATGACTTG	GTTATCACAA	AAACTGGTTG	TCAAGTCTTG
1090	1100	1110	1120	1130	1140
ACCTTGGCAC	CCAAGAATT	AATTGTATTG	TAA.....	.....	.....

(SEQ ID No: 33 &amp; 34)

**FIG. 4A**

5'		5855	5864	5873	5882	5891	5900
	ATG CTA CAA TAT TCT CAC AAG TTA CCA AAG GAG TTC GCG ATG TCA GGA TTT TTA	M L Q Y S Q K L P K E F A M S G F L					
	GAA CAA CGA TTA GGT CAC TGC CTA AGG CAG ATG GCA GAG AAG GGG CTA GAG GCT	E Q R L G H C L R Q M A E K G L E A					
	CTT CTA GTC ACC CAT TTA ACC AAT AGT TAT TAC TTG ACA GGT TTT TCT GGA ACT	L L V T H L L N S Y Y L T G F S G T					
	GCA GCA ACT GTT TTG ATA ACG GCC AAA CGT CGT GTT TTG ATC ACA GAT TCA CGT	A A T V L I T A K R R V R V L I T D S R					
	TAT ACC TTG CTT GCT AAA GCT AGT GTT GAG GGA TTT GAT ATT ATC GAA AGC CGC	Y T L L A K A S V E G F D I I E S R					
	ACG CCG CTT AAG GTT GTG GCA GAA TTG TTA GAG GCT GAT CAA ATA GAT TGC CTT	T P L K V V A E L L L E A D Q I D C L					



6/18

FIG. 4B

6179	6188	6197	6206	6215	6224
GGT TTT GAG GAC CAG GTA TCG TTT TCT TTT TAC CAG GCC ATG CAA GCA GAA CTG					
G F E D Q V S F S F F Y Q A M Q A E L					
6233	6242	6251	6260	6269	6278
TCA GGA ATA ACC TTG CTT GCT CAG TCA GGT TTT GTG GAG CAT TTA CGT CTT ATT					
S G I T L L A Q S G F V E H L R L I					
6287	6296	6305	6314	6323	6332
AAG GAC GCC TCT GAA ATC GAT ACC ATT GCT AAA GCG TGC TCG ATC TCA GAC AAA					
K D A S E I D T I A K A C S I S D K					
6341	6350	6359	6368	6377	6386
GCA TTT GAA GAT GCT CTT GAT TTT ATT AAA CCA GGG ACA ACC ACT GAA CGT GAC					
A F E D A L D F I K P G T T T E R D					
6395	6404	6413	6422	6431	6440
CTG GCT AAT TTT TTA GAT TTT CGT ATG CGT CAG TAT GGT GCC AGC GGC ACA TCA					
L A N F L D F R M R Q Y G A S G T S					
6449	6458	6467	6476	6485	6494
TTT GAT ATC ATT GTA GCT TCA GGC TAT CTC TCT GCC ATG CCT CAT GGA CGC GCC					
F D I I V A S S G Y L S A M P H G R A					

(SEQ ID No: 34)

7/18

FIG. 4C

6503	AGT GAC AAG	6512	GTT ATC CAG	6521	AAA GAG AGC	6530	TTG ACC ATG	6539	GAC TTT GGG	6548	TGT TAC
---	---	---	---	---	---	---	---	---	---	---	---
S	D K V	I Q	N K E	S L	T M	D F	G C	Y			
6557	TAC AAT CAC	6566	TAT GTT AGT	6575	ATG ACG AGG	6584	ATT ACC CAT	6593	ATT GGC CAA	6602	GTT ACT
---	---	---	---	---	---	---	---	---	---	---	---
Y	N H	Y S	V Y	M T	R T	I H	I G	Q V			
6611	GAT GAA GAA	6620	CGT GAG ATT	6629	GCT CTT GTT	6638	GCT GCT CTT	6647	AAT AAG GCT	6656	TTA ATT
---	---	---	---	---	---	---	---	---	---	---	---
D	E E R	E I	A Y	L A	V L	A A	N K	A L			
6665	GCT AAA GCT	6674	AGC GCT GGC	6683	ATG ACT TAT	6692	TTT GAC GGT	6701	ATT CCG CGC	6710	CAA
---	---	---	---	---	---	---	---	---	---	---	---
A	K A S	A G	M T	Y S	D F	D G	I P	R Q			
6719	CTC ATC ACT	6728	GAG GCG GGT	6737	GGC AGT CGC	6746	ACA CAT GGC	6755	ATT GGT CAT	6764	GGC
---	---	---	---	---	---	---	---	---	---	---	---
L	I T E	A G	Y G	S S	R F	H T	G I	G H			

8/18

FIG. 4D

6773	ATC	GGG	CTT	GAC	ATC	CAT	6782	GAG	AAT	CCA	TTT	GGG	AAA	TCT	GAG	CAA	CTT	6818	CTC
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
I	G	L	D	I	H	E	N	P	F	F	G	K	S	E	Q	L	L	---	---
6827	CAA	GCT	GGA	ATG	GTG	GTA	ACA	GAT	GAG	CCA	GGT	ATC	TAT	TTG	GAT	AAC	AAA	6872	TAT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Q	A	G	M	V	V	T	D	E	P	G	I	Y	L	D	N	K	Y	---	---
6881	GGT	GTC	CGT	ATT	GAA	GAT	6890	GAC	TTG	GTT	ATC	ACA	AAA	GGT	TGT	CAA	GTC	6926	TTG
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
G	V	R	I	E	D	D	L	V	I	T	K	ACT	GGT	G	C	Q	V	---	---
6935	ACC	TTG	GCA	CCC	AAA	GAA	6944	TTA	ATT	GTA	TTG	TAA	3'	---	---	---	---	---	---
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
T	L	A	P	K	E	L	I	V	L	*	---	---	---	---	---	---	---	---	---

9/18

FIG. 5

ORIGIN

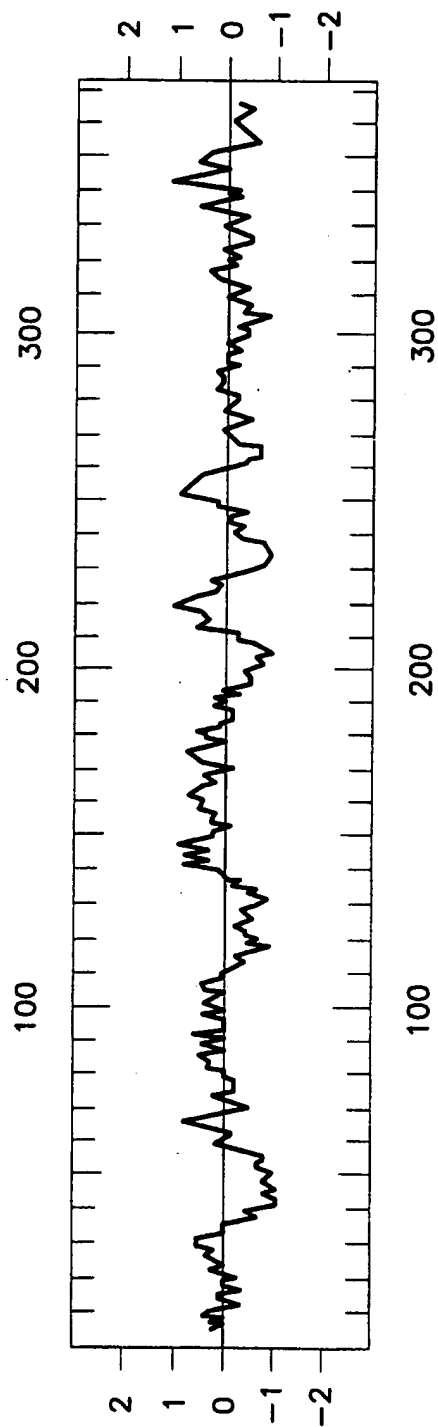
1	mkleklrnl	gqlgidgmli	tsntnvm	qftgsaglav	isgdkaafit	dfryteqakv
61	qvkgfeileh	ggsliqttad	tvesfgikrl	gfeqnsmtyg	tyasysavis	daelpvpaes
121	veklrlkss	eeikileeaa	kiaddafrhi	ltfmkpgise	lavanelefy	mrsqgadsss
181	fdmivasglr	sslpghvasd	kliesgdltv	ldfgayykgy	csditrtvav	gqpsdqlkei
241	yqvfvdaqal	gvahikpgmt	gkeadaltrd	hiaakgygdy	fghstghglg	mevhespqls
301	vrssailepg	mvvtvepgiy	ipetggvrie	ddivitengn	rtithspkel	ii

(SEQ ID No: 3)

//

10/18

FIG. 6



11/18

FIG. 7

ORIGIN

1	mriecklkvkm	ltenidslli	tdmknifylt	gfsqtagtvm	ltqkrnifmt	dsrysemarg
61	liknfeilet	rdpislltel	sasesvknma	feetvdyaff	krkskaatkl	dlfstnsfvl
121	elrqikdese	islkkacei	adeafmsalr	fiexpgrtele	vanfldfkmr	dleasgisfe
181	tivasgkrss	lphgvatskm	iqfgdpvtid	fgcyyehyas	dmtrtifvgs	vddkmrtiye
241	tvrkaneali	kqvkagmtya	qydniprevi	ekadfgqyft	hgighglgld	vheipyfnqs
301	mtenqlrsgm	vitdepgiyl	pefggvried	dllvtengce	vltkapkeli	vi

(SEQ ID No: 4)

//

12/18

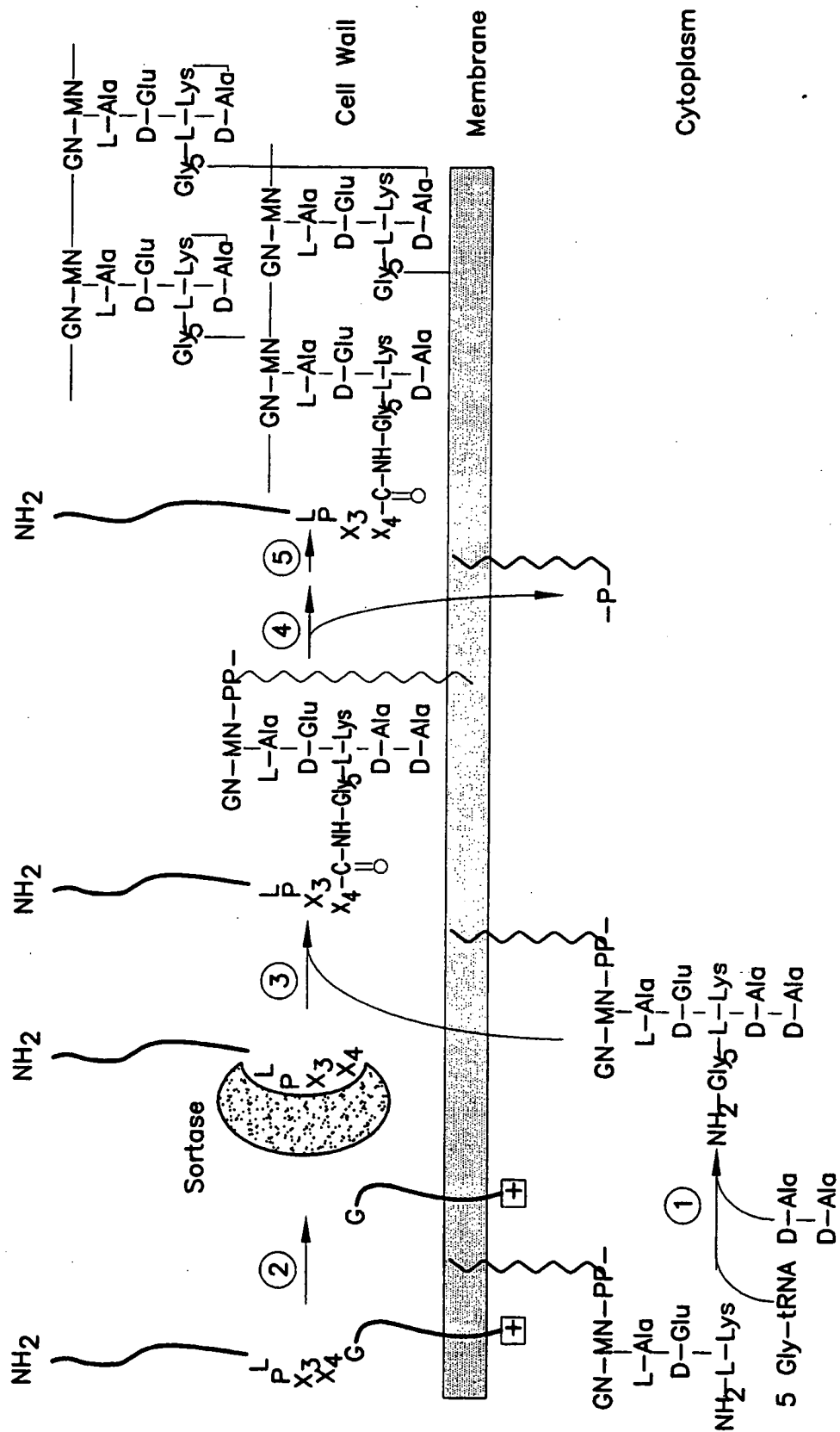
FIG. 8

## ORIGIN

1	mnldklqnwl	qengmdvayv	sspttynyft	gfildpeeri	fkfakfdae	pflfcpalny
61	eeakasawdg	dvvgylsds	pwskliaeeik	krtkdyqnwa	veknngltvah	yqalhaqfpg
121	sdfskdlsdf	iahirlfkte	selvklrkag	eeadfafgig	fealrngvte	ravvsqieyq
181	lklqkgvmqt	sfdtqvqagk	naanphqggs	mntvqpnelv	lfdlgtmhg	yasdssrtva
241	ygeptdkmre	iyevnrtaqq	aaldaakpgm	taselldgvar	kiitdaggye	yfihrhghgi
301	gmevhefpsi	angndvvlee	gmcfsiepgi	yipgfagvri	edcgvltkdg	fkpfthtske
361	lkvlpvke					

(SEQ. ID No: 5)

FIG. 9





14/18

FIG. 10A

```

10      20      30      40      50      60
--//--CGACTC TAGAGGACAA GCAACTAAGC AGGCGCCAAA TTATGAAATT ATTAATCGTA

70      80      90      100     110     120
AATCTACTAT TATTGGTGAG ATTAAAAAAT TGCTACNCCA ANAAAAATTTT GAAAAATGTTN

130     140     150     160     170     180
GTTTTGAGGG GCATCATGTN NTTTATGATA CATACCTTGA ATTAAATAAA AGCCGTATAT

190     200     210     220     230     240
CATTAAATAAG CATTTCCTAAT ACTGTAGATA AAATTAGAGA CGTCCAAGAT GCTGACGAAA

250     260     270     280     290     300
TTGCTTTTAAT TCAAAAAGCA GCTAATATTG TTGATGPAAC ATATGAATAT ATTTTAACTG

310     320     330     340     350     360
TTGTAAAAGC NNGCATGACT GAAAAANAAT TAAAGGCAAT ATTAGAAAGC CAAATGCTAN

370     380     390     400     410     420
AATTAGGAGC AGATGGACCN TCTTTCGATA CGATTGTNGC ATCTGGTCCT AGAGGTGCAT

430     440     450     460     470     480
TACCACCATG GTGTTGCAAG TGATAAATTT ATTGAAAAAG GCGACATGAT TACATTANAT

490     500     510     520     530     540
TTTGGGGCCN TTNTTATAAC CGGCTATTGT TCAAATTTTA CTANAACATT TTGCNATTGG

```

15/18

FIG. 10B

```

550      560      570      580      590      600
GAAAAACCG ATCCNTAAC TGAAGAAGAA-  // -ATTGTAG CATCTGTTCA TAGAGGTGCA

610      620      630      640      650      660
TTACCACATG GTGTTGCAAG TGATAAAATT ATTGAAAAAG GCGACATGAT NACATTAGAT

670      680      690      700      710      720
TTNGGCGCGT ATTATAACGG CTATTGTTCA GATATTACTA GAACATTTCG TATTGGAGAA

730      740      750      760      770      780
CCAGATCCTA AACTGAAAGA AATATATCAA ATAGTACTTG AATCTCAAT GAAAGCAATT

790      800      810      820      830      840
AATGAGATTA GACCTGGCAT GACTGGTGCA GAAGCTGATG CCATTTCAG AACTATTTA

850      860      870      880      890      900
GAGTCAAAAG GGTATGGAAA AGAATTGGA CATTCACTAG GACATGGTAT TGGTTTAGAA

910      920      930      940      950      960
ATCCATGAAG GGCCATGCT GGNTCGTAGC ATACAAGATA AACTTCAAGT TAACAACGTGT

970      980      990      1000      1010      1020
GTTACAGTAG NACCTGGTGT TNATATAGAA GGTTGGGCG GTATAAGAA AGAAGAGATA

1030      1040      1050      1060      1070      1080
TTTTAA.... ..... 1050 ..... 1060 ..... 1070 ..... 1080

```

(SEQ ID No: 28 &amp; 29)

16/18

FIG. 11

5'	GAT	CCT	AAA	CTG	AAA	GAA	ATA	TAT	CAA	ATA	GTA	CTT	GAA	TCT	CAA	ATG	AAA	GCA	54
	D	P	K	L	K	E	I	Y	Q	I	V	L	E	S	Q	M	K	A	
	ATT	AAT	GAG	ATT	AGA	CCT	GGC	ATG	ACT	GGT	GCA	GAA	GCT	GAT	GCC	ATT	TCA	AGA	108
	I	N	E	I	R	P	G	M	T	G	A	E	A	D	A	I	S	R	
	AAC	TAT	TTA	GAG	TCA	AAA	GGG	TAT	GGA	AAA	GAA	TTT	GGA	CAT	TCA	CTA	GGA	CAT	162
	N	Y	L	E	S	K	G	Y	G	K	E	F	G	H	S	L	G	H	
	GGT	ATT	GGT	TTA	GAA	ATC	CAT	GAA	GGG	CCA	ATG	CTG	GCT	CGT	ACG	ATA	CAA	GAT	216
	G	I	G	L	E	I	H	E	G	P	M	L	A	R	T	I	Q	D	
	AAA	CTT	CAA	GTT	AAC	AAC	TGT	GTT	ACA	GTA	GAA	CCT	GGT	GTT	TAT	ATA	GAA	GGT	270
	K	L	Q	V	N	N	C	V	T	V	E	P	G	V	Y	I	E	G	
	TTG	GGC	GGT	ATA	AGA	ATA	GAA	GAT	GAT	ATT	TTA	ATT	ACA	GAA	AAT	GGT	TGT	CAA	324
	L	G	G	I	R	I	E	D	D	I	L	I	T	E	N	G	C	Q	
	GTC	TTT	ACT	AAA	TGC	ACA	AAA	GAC	CTT	ATA	GTT	TTA	ACA	TAA	3'				
	V	F	T	K	C	T	K	D	L	I	V	L	T	*					

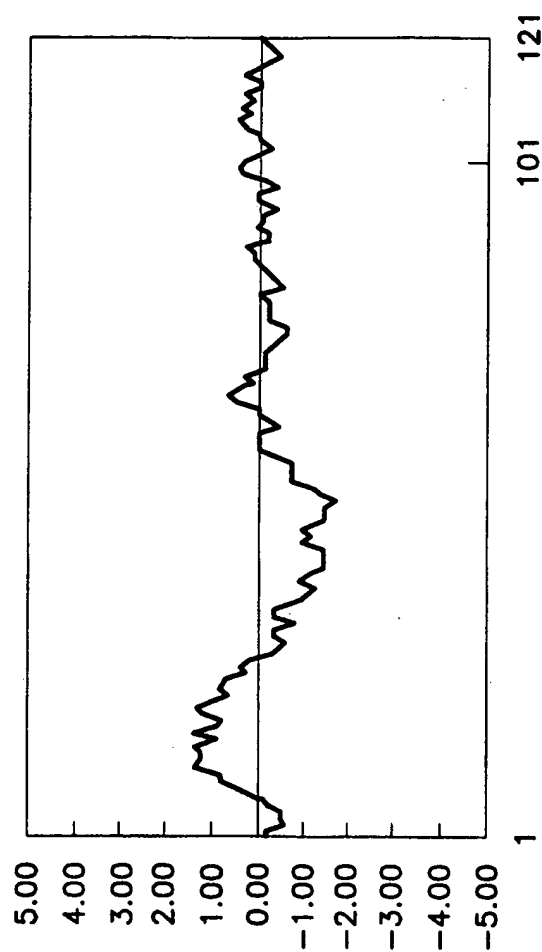
(SEQ ID No: 2)

**FIG. 12**

5'		3266		3275		3284		3293		3302		3311
	ATG	GTC AAA	GTA	ACT GAT	TAT	TCA AAT	TCA	AAA TTA	GGT	AAA GTA	GAA	ATA GCG
	---	---	---	---	---	---	---	---	---	---	---	---
	Met	Val Lys	Val	Thr Asp	Tyr Ser	Ser Asn	Ser	Lys Leu	Gly	Lys Val	Glu Ile	Ala Ala
		3320		3329		3338		3347		3356		3365
	CCA	GAA GTG	CTA	TCT GTT	ATT	GCA AGT	ATA	GCT ACT	TCG	GAA GTC	GAA	GGC ATC
	---	---	---	---	---	---	---	---	---	---	---	---
	Pro	Glu Val	Leu Ser	Val Ile	Ile Ala	Ser Ser	Ile Thr	Ser		Glu Val	Glu Gly	Ile Ile
		3374		3383		3392		3401		3410		3419
	ACT	GGC CAT	TTT	GCT GAA	TTA	AAA GAA	ACA	AAT TTA	GAA	AAA GTT	AGT	CGT AAA
	---	---	---	---	---	---	---	---	---	---	---	---
	Thr	Gly His	Phe	Ala Glu	Leu Leu	Lys Glu	Thr Asn	Leu Glu	Lys	Lys Val	Ser Arg	Lys Lys
		3428		3437		3446		3455		3464		3473
	AAT	TTA AGC	CGT	GAT TTA	AAA	ATC GAG	AGT	AAA GAA	GAT	GGC ATA	TAT	ATA GAT
	---	---	---	---	---	---	---	---	---	---	---	---
	Asn	Leu Ser	Arg Asp	Leu Leu	Lys Ile	Glu Ser	Lys Glu	Asp	Gly Ile	Tyr Ile	Asp	
		3482		3491		3500		3509		3518		3527
	GTA	TAT TGT	GCA	TTA AAA	CAT	GGT GTT	AAT	ATT TCA	AAA	ACT GCA	AAC	AAA ATT
	---	---	---	---	---	---	---	---	---	---	---	---
	Val	Tyr Cys	Ala Leu	Lys Lys	His Gly	Val Val	Asn Ile	Ser Lys	Thr Ala	Asn	Lys Ile	
		3536		3545		3554		3563		3572		3581
	CAA	ACG TCA	ATT	TTT AAT	TCA	ATT TCT	AAT	ATG ACA	GCG	ATA GAA	CCT	AAG CAA
	---	---	---	---	---	---	---	---	---	---	---	---
	Gln	Thr Ser	Ile Phe	Asn Ser	Ile Ser	Ile Ser	Asn Met	Thr Ala	Ile Ile	Glu Pro	Lys Lys	Gln Gln
		3590		3599		3608		3617				
	ATT	AAT ATT	CAC	ATT ACA	CAA	ATC GTT	ATT	GAA AAG	TAA	3'		
	---	---	---	---	---	---	---	---	---			
	Ile	Asn Ile	His Ile	Thr Thr	Gln Ile	Val Ile	Ile Glu	Lys	***			

(SEQ ID No: 30 & 31)

FIG. 13



-1-

## SEQUENCE LISTING

5           <110> Schneewind, Olaf  
            Ton-That, Hung  
            Mazmanian, Sarkis

            <120> IDENTIFICATION OF SORTASE GENE

10           <130> 30435.46W001

            <150> 60/055,662  
            <151> 1997-08-14

15           <160> 34

            <170> FastSEQ for Windows Version 3.0

20           <210> 1  
            <211> 5  
            <212> PRT  
            <213> Artificial Sequence

25           <220>  
            <223> Mutated derived from Streptococcus Pyogenes

            <400> 1  
30      Leu Pro Xaa Thr Gly  
          1                   5

            <210> 2  
            <211> 121  
            <212> PRT

35           <213> Staphylococcus Aureus

            <400> 2  
40      Asp Pro Lys Leu Lys Glu Ile Tyr Gln Ile Val Leu Glu Ser Gln Met  
          1           5           10           15  
40      Lys Ala Ile Asn Glu Ile Arg Pro Gly Met Thr Gly Ala Glu Ala Asp  
          20           25           30  
40      Ala Ile Ser Arg Asn Tyr Leu Glu Ser Lys Gly Tyr Gly Lys Glu Phe  
          35           40           45  
45      Gly His Ser Leu Gly His Gly Ile Gly Leu Glu Ile His Glu Gly Pro  
          50           55           60  
45      Met Leu Ala Arg Thr Ile Gln Asp Lys Leu Gln Val Asn Asn Cys Val  
          65           70           75           80  
45      Thr Val Glu Pro Gly Val Tyr Ile Glu Gly Leu Gly Gly Ile Arg Ile  
          85           90           95  
50      Glu Asp Asp Ile Leu Ile Thr Glu Asn Gly Cys Gln Val Phe Thr Lys  
          100           105           110  
50      Cys Thr Lys Asp Leu Ile Val Leu Thr  
          115           120

55           <210> 3  
            <211> 353  
            <212> PRT  
            <213> Bacillus Sp.

60           <400> 3  
60      Met Lys Leu Glu Lys Leu Arg Asn Leu Phe Gly Gln Leu Gly Ile Asp  
          1           5           10           15  
60      Gly Met Leu Ile Thr Ser Asn Thr Asn Val Arg Val Met Thr Gly Phe  
          20           25           30  
65      Thr Gly Ser Ala Gly Leu Ala Val Ile Ser Gly Asp Lys Ala Ala Phe  
          35           40           45  
65      Ile Thr Asp Phe Arg Tyr Thr Glu Gln Ala Lys Val Gln Val Lys Gly  
          50           55           60  
65      Phe Glu Ile Ile Glu His Gly Gly Ser Leu Ile Gln Thr Thr Ala Asp



-3-

195 200 205  
 Ile Asp Phe Gly Cys Tyr Tyr Glu His Tyr Ala Ser Asp Met Thr Arg  
 210 215 220  
 5 Thr Ile Phe Val Gly Ser Val Asp Asp Lys Met Arg Thr Ile Tyr Glu  
 225 230 235 240  
 Thr Val Arg Lys Ala Asn Glu Ala Leu Ile Lys Gln Val Lys Ala Gly  
 245 250 255  
 Met Thr Tyr Ala Gln Tyr Asp Asn Ile Pro Arg Glu Val Ile Glu Lys  
 260 265 270  
 10 Ala Asp Phe Gly Gln Tyr Phe Thr His Gly Ile Gly His Gly Leu Gly  
 275 280 285  
 Leu Asp Val His Glu Ile Pro Tyr Phe Asn Gln Ser Met Thr Glu Asn  
 290 295 300  
 15 Gln Leu Arg Ser Gly Met Val Ile Thr Asp Glu Pro Gly Ile Tyr Leu  
 305 310 315 320  
 Pro Glu Phe Gly Gly Val Arg Ile Glu Asp Asp Leu Leu Val Thr Glu  
 325 330 335  
 Asn Gly Cys Glu Val Leu Thr Lys Ala Pro Lys Glu Leu Ile Val Ile  
 340 345 350  
 20  
 <210> 5  
 <211> 368  
 <212> PRT  
 <213> Lactobacillus Delbrueckii Lactis  
 25  
 <400> 5  
 Met Asn Leu Asp Lys Leu Gln Asn Trp Leu Gln Glu Asn Gly Met Asp  
 1 5 10 15  
 30 Val Ala Tyr Val Ser Ser Pro Thr Thr Ile Asn Tyr Phe Thr Gly Phe  
 20 25 30  
 Ile Thr Asp Pro Glu Glu Arg Ile Phe Lys Leu Phe Ala Phe Lys Asp  
 35 40 45  
 Ala Glu Pro Phe Leu Phe Cys Pro Ala Leu Asn Tyr Glu Glu Ala Lys  
 50 55 60  
 35 Ala Ser Ala Trp Asp Gly Asp Val Val Gly Tyr Leu Asp Ser Glu Asp  
 65 70 75 80  
 Pro Trp Ser Lys Ile Ala Glu Glu Ile Lys Lys Arg Thr Lys Asp Tyr  
 85 90 95  
 40 Gln Asn Trp Ala Val Glu Lys Asn Gly Leu Thr Val Ala His Tyr Gln  
 100 105 110  
 Ala Leu His Ala Gln Phe Pro Asp Ser Asp Phe Ser Lys Asp Leu Ser  
 115 120 125  
 Asp Phe Ile Ala His Ile Arg Leu Phe Lys Thr Glu Ser Glu Leu Val  
 130 135 140  
 45 Lys Leu Arg Lys Ala Gly Glu Glu Ala Asp Phe Ala Phe Gln Ile Gly  
 145 150 155 160  
 Phe Glu Ala Leu Arg Asn Gly Val Thr Glu Arg Ala Val Val Ser Gln  
 165 170 175  
 Ile Glu Tyr Gln Leu Lys Leu Gln Lys Gly Val Met Gln Thr Ser Phe  
 180 185 190  
 50 Asp Thr Ile Val Gln Ala Gly Lys Asn Ala Ala Asn Pro His Gln Gly  
 195 200 205  
 Pro Ser Met Asn Thr Val Gln Pro Asn Glu Leu Val Leu Phe Asp Leu  
 210 215 220  
 55 Gly Thr Met His Glu Gly Tyr Ala Ser Asp Ser Ser Arg Thr Val Ala  
 225 230 235 240  
 Tyr Gly Glu Pro Thr Asp Lys Met Arg Glu Ile Tyr Glu Val Asn Arg  
 245 250 255  
 60 Thr Ala Gln Gln Ala Ala Ile Asp Ala Ala Lys Pro Gly Met Thr Ala  
 260 265 270  
 Ser Glu Leu Asp Gly Val Ala Arg Lys Ile Ile Thr Asp Ala Gly Tyr  
 275 280 285  
 Gly Glu Tyr Phe Ile His Arg Leu Gly His Gly Ile Gly Met Glu Val  
 290 295 300  
 65 His Glu Phe Pro Ser Ile Ala Asn Gly Asn Asp Val Val Leu Glu Glu  
 305 310 315 320  
 Gly Met Cys Phe Ser Ile Glu Pro Gly Ile Tyr Ile Pro Gly Phe Ala  
 325 330 335  
 Gly Val Arg Ile Glu Asp Cys Gly Val Leu Thr Lys Asp Gly Phe Lys



-4-

	340	345	350
	Pro Phe Thr 355	His Thr Ser Lys Glu Leu Lys Val Leu Pro Val Lys Glu 360                365	
5	<210>	6	
	<211>	25	
	<212>	PRT	
	<213>	Staphylococcus Aureus	
10	<400>	6	
	Glu Glu Asn      Pro Phe Ile Gly Thr Thr      Val Phe Gly Gly Leu Ser Leu 1                                5                                10                                15		
	Ala Leu Gly      Ala Ala Leu Leu Ala Gly 20                                25		
15	<210>	7	
	<211>	23	
	<212>	PRT	
	<213>	Staphylococcus Aureus	
20	<400>	7	
	Gly Glu Glu Ser Thr Asn Lys Gly Met      Leu Phe Gly Gly Leu Phe Ser 1                                5                                10                                15		
	Ile Leu Gly      Leu Ala Leu Leu 20		
25	<210>	8	
	<211>	24	
	<212>	PRT	
	<213>	Staphylococcus Sobrinos	
30	<400>	8	
	Asp Ser Ser Asn Ala Tyr Leu Pro Leu      Leu Gly Leu Val Ser Leu Thr 1                                5                                10                                15		
	Ala Gly Phe      Ser Leu Leu Gly Leu 20		
35	<210>	9	
	<211>	24	
	<212>	PRT	
	<213>	Enterococcus Faecalis	
40	<400>	9	
	Glu Lys Gln Asn Val Leu Leu Thr Val      Val Gly Ser Leu Ala Ala Met 1                                5                                10                                15		
	Leu Gly Leu      Ala Gly Leu Gly Phe 20		
45	<210>	10	
	<211>	23	
	<212>	PRT	
	<213>	Streptococcus Pyogenes	
50	<400>	10	
	Ser Ile Gly Thr Tyr Leu Phe Lys Ile      Gly Ser Ala Ala Met Ile Gly 1                                5                                10                                15		
	Ala Ile Gly      Ile Tyr Ile Val 20		
55	<210>	11	
	<211>	22	
	<212>	PRT	
	<213>	Listeria Monocytogenes	
60	<400>	11	
	Asp Ser Asp Asn Ala Leu Tyr Leu Leu      Leu Gly Leu Leu Ala Val Gly 1                                5                                10                                15		
	Thr Ala Met      Ala Leu Thr 20		

-5-

5           <210> 12  
            <211> 5  
            <212> PRT  
            <213> Staphylococcus Aureus

10           <400> 12  
            Arg Arg Arg Glu Leu  
            1                   5

15           <210> 13  
            <211> 9  
            <212> PRT  
            <213> Staphylococcus Aureus

20           <400> 13  
            Arg Arg Asn Lys Lys Asn His Lys Ala  
            1                   5

25           <210> 14  
            <211> 5  
            <212> PRT  
            <213> Staphylococcus Sobrinus

30           <400> 14  
            Arg Arg Lys Gln Asp  
            1                   5

35           <210> 15  
            <211> 7  
            <212> PRT  
            <213> Enterococcus Faecalis

40           <400> 15  
            Lys Arg Arg Lys Glu Thr Lys  
            1                   5

45           <210> 16  
            <211> 5  
            <212> PRT  
            <213> Streptococcus Pyogenes

50           <400> 16  
            Lys Arg Arg Lys Ala  
            1                   5

55           <210> 17  
            <211> 8  
            <212> PRT  
            <213> Actinomyces Viscosus

60           <400> 17  
            Lys Arg Arg His Val Ala Lys His  
            1                   5

65           <210> 18  
            <211> 5  
            <212> PRT  
            <213> Streptococcus Agalactiae

70           <400> 18  
            Lys Arg Arg Lys Ser  
            1                   5

75           <210> 19  
            <211> 6  
            <212> PRT  
            <213> Streptococcus Pyogenes

-6-

<400> 19  
Lys Arg Lys Glu Glu Asn  
1 5

5 <210> 20  
<211> 5  
<212> PRT  
<213> Artificial Sequence

10 <220>  
<223> Mutated derived from Streptococcus Pyogenes

<400> 20  
Arg Arg Arg Glu Ser  
15 1 5

<210> 21  
<211> 5  
<212> PRT  
20 <213> Artificial Sequence

<220>  
<223> Mutated derived from Streptococcus Pyogenes

25 <400> 21  
Arg Arg Arg Ser Leu  
1 5

<210> 22  
<211> 5  
<212> PRT  
30 <213> Artificial Sequence

<220>  
35 <223> Mutated derived from Streptococcus Pyogenes

<400> 22  
Arg Arg Ser Glu Leu  
40 1 5

<210> 23  
<211> 5  
<212> PRT  
45 <213> Artificial Sequence

<220>  
<223> Mutated derived from Streptococcus Pyogenes

<400> 23  
50 Arg Ser Arg Glu Leu  
1 5

<210> 24  
<211> 5  
<212> PRT  
55 <213> Artificial Sequence

<220>  
<223> Mutated derived from Streptococcus Pyogenes

60 <400> 24  
Ser Arg Arg Glu Leu  
1 5

65 <210> 25  
<211> 5  
<212> PRT  
<213> Artificial Sequence

-7-

```

<220>
<223> Mutated derived from Streptococcus Pyogenes

<400> 25
5   Arg Arg Ser Ser Ser
    1           5

    <210> 26
    <211> 5
    <212> PRT
    <213> Artificial Sequence

    <220>
    <223> Mutated derived from Streptococcus Pyogenes
15

    <400> 26
    Arg Ser Arg Ser Ser
    1           5

    <210> 27
    <211> 5
    <212> PRT
    <213> Artificial Sequence

    <220>
    <223> Mutated derived from Streptococcus Pyogenes
25

    <400> 27
    Ser Arg Arg Ser Ser
    1           5

    <210> 28
    <211> 1017
    <212> DNA
    <213> Staphylococcus Aureus

    <220>
    <221> CDS
    <222> (716)...(1014)
40

    <400> 28
cgactctaga ggacaagcaa ctaagcaggc gccaaattat gaaattatta atcgtaaadc 60
tactattatt ggtgagatta aaaaattgct acnccaanaa aattttgaaa atgttngttt 120
tgaggggcat catgtntttt atgatacata ccttgaatta aataaaaagcc gtatatcatt 180
45 aataagcatt tctaatactg tagataaaat tagagacgtc caagatgctg acgaaattgc 240
ttaaattcaa aaagcagcta atattgttga tgaacacata gaatatattt taactgttgt 300
aaaagcnnngc atgactgaaa aanaattaaa ggcaatatta gaaagccaaa tgctanaaatt 360
aggagcagat ggacctctt tcgatacgat tgtngcatct ggtcctagag gtgcattacc 420
accatggtgt tgcaagtgat aaaattattg aaaaaggcga catgattaca ttanattttg 480
50 ggcgcntntt tataaccggc tattgttcaa attttactan aacattttgc nattgggaaa 540
aaccgatcc ntaactgaa aagaaattgt agcatctgtt catagagggt cattaccaca 600
tgggtgttga agtgataaaa ttattgaaaa aggcgacatg atnacattag atttnggcgc 660
gtattataac ggctattgtt cagatattac tagaacattt gctattggag aacca gat 718
                                Asp
                                1

cct aaa ctg aaa gaa ata tat caa ata gta ctt gaa tct caa atg aaa 766
Pro Lys Leu Lys Glu Ile Tyr Gln Ile Val Leu Glu Ser Gln Met Lys
                                5
                                10
                                15

60 gca att aat gag att aga cct ggc atg act ggt gca gaa gct gat gcc 814
Ala Ile Asn Glu Ile Arg Pro Gly Met Thr Gly Ala Ala Asp Ala
                                20
                                25
                                30

65 att tca aga aac tat tta gag tca aaa ggg tat gga aaa gaa ttt gga 862
Ile Ser Arg Asn Tyr Leu Glu Ser Lys Gly Tyr Gly Lys Glu Phe Gly
                                35
                                40
                                45

cat tca cta gga cat ggt att ggt tta gaa atc cat gaa ggg cca atg 910

```

-8-

His Ser Leu Gly His Gly Ile Gly Leu Glu Ile His Glu Gly Pro Met  
 50 55 60 65  
 5 ctg gnt cgt acg ata caa gat aaa ctt caa gtt aac aac tgt gtt aca 958  
 Leu Xaa Arg Thr Ile Gln Asp Lys Leu Gln Val Asn Asn Cys Val Thr  
 70 75 80  
 10 gta gna cct ggt gtt nat ata gaa ggt ttg ggc ggt ata aga ata gaa 1006  
 Val Xaa Pro Gly Val Xaa Ile Glu Gly Leu Gly Gly Ile Arg Ile Glu  
 85 90 95  
 gag ata tt taa 1017  
 Glu Ile  
 15  
 <210> 29  
 <211> 100  
 <212> PRT  
 20 <213> Staphylococcus Aureus  
 <400> 29  
 Asp Pro Lys Leu Lys Glu Ile Tyr Gln Ile Val Leu Glu Ser Gln Met  
 1 5 10 15  
 25 Lys Ala Ile Asn Glu Ile Arg Pro Gly Met Thr Gly Ala Glu Ala Asp  
 20 25 30  
 Ala Ile Ser Arg Asn Tyr Leu Glu Ser Lys Gly Tyr Gly Lys Glu Phe  
 35 40 45  
 30 Gly His Ser Leu Gly His Gly Ile Gly Leu Glu Ile His Glu Gly Pro  
 50 55 60  
 Met Leu Xaa Arg Thr Ile Gln Asp Lys Leu Gln Val Asn Asn Cys Val  
 65 70 75 80  
 Thr Val Xaa Pro Gly Val Xaa Ile Glu Gly Leu Gly Gly Ile Arg Ile  
 85 90 95  
 35 Glu Glu Ile Phe  
 100  
 <210> 30  
 <211> 360  
 40 <212> DNA  
 <213> Staphylococcus Aureus  
 <220>  
 <221> CDS  
 45 <222> (1)... (357)  
 <400> 30  
 atg gtc aaa gta act gat tat tca aat tca aaa tta ggt aaa gta gaa 48  
 Met Val Lys Val Thr Asp Tyr Ser Asn Ser Lys Leu Gly Lys Val Glu  
 1 5 10 15  
 ata gcg cca gaa gtg cta tct gtt att gca agt ata gct act tcg gaa 96  
 Ile Ala Pro Glu Val Leu Ser Val Ile Ala Ser Ile Ala Thr Ser Glu  
 20 25 30  
 55 gtc gaa ggc atc act ggc cat ttt gct gaa tta aaa gaa aca aat tta 144  
 Val Glu Gly Ile Thr Gly His Phe Ala Glu Leu Lys Glu Thr Asn Leu  
 35 40 45  
 60 gaa aaa gtt agt cgt aaa aat tta agc cgt gat tta aaa atc gag agt 192  
 Glu Lys Val Ser Arg Lys Asn Leu Ser Arg Asp Leu Lys Ile Glu Ser  
 50 55 60  
 65 aaa gaa gat ggc ata tat ata gat gta tat tgt gca tta aaa cat ggt 240  
 Lys Glu Asp Gly Ile Tyr Ile Asp Val Tyr Cys Ala Leu Lys His Gly  
 65 70 75 80  
 aat att tca aaa act gca aac aaa att caa acg tca att ttt aat tca 288  
 Asn Ile Ser Lys Thr Ala Asn Lys Ile Gln Thr Ser Ile Phe Asn Ser

-9-

	85	90	95	
5	att tct aat atg aca gcg ata gaa cct aag caa att aat att cac att Ile Ser Asn Met Thr Ala Ile Glu Pro Lys Gln Ile Asn Ile His Ile 100 105 110			336
10	aca caa atc gtt att gaa aag taa Thr Gln Ile Val Ile Glu Lys 115			360
15	<210> 31 <211> 119 <212> PRT <213> Staphylococcus Aureus			
20	<400> 31 Met Val Lys Val Thr Asp Tyr Ser Asn Ser Lys Leu Gly Lys Val Glu 1 5 10 15 Ile Ala Pro Glu Val Leu Ser Val Ile Ala Ser Ile Ala Thr Ser Glu 20 25 30 Val Glu Gly Ile Thr Gly His Phe Ala Glu Leu Lys Glu Thr Asn Leu 35 40 45 Glu Lys Val Ser Arg Lys Asn Leu Ser Arg Asp Leu Lys Ile Glu Ser 50 55 60 Lys Glu Asp Gly Ile Tyr Ile Asp Val Tyr Cys Ala Leu Lys His Gly 65 70 75 80 Asn Ile Ser Lys Thr Ala Asn Lys Ile Gln Thr Ser Ile Phe Asn Ser 85 90 95 Ile Ser Asn Met Thr Ala Ile Glu Pro Lys Gln Ile Asn Ile His Ile 100 105 110 Thr Gln Ile Val Ile Glu Lys 115			
35	<210> 32 <211> 19 <212> PRT <213> Artificial Sequence			
40	<220> <223> Mutated derived from Streptococcus Pyogenes			
45	<400> 32 His His His His His Ala Gln Ala Leu Glu Pro Thr Gly Glu Glu 1 5 10 15 Asn Pro Phe			
50	<210> 33 <211> 1119 <212> DNA <213> Streptococcus Pyogenes			
55	<220> <221> CDS <222> (1)...(995)			
60	<400> 33 atg cta caa tat tct caa aag tta cca aag gag ttc gcg atg tca gga Met Leu Gln Tyr Ser Gln Lys Leu Pro Lys Glu Phe Ala Met Ser Gly 1 5 10 15			48
65	ttt tta gaa caa cga tta ggt cac tgc cta agg cag atg gca gag aag Phe Leu Glu Gln Arg Leu Gly His Cys Leu Arg Gln Met Ala Glu Lys 20 25 30			96
	ggg cta gag gct ctt cta gtc acc cat tta acc aat agt tat tac ttg Gly Leu Glu Ala Leu Leu Val Thr His Leu Thr Asn Ser Tyr Tyr Leu 35 40 45			144

-10-

	aca ggt ttt tct gga act gca gca act gtt ttg ata acg gcc aaa cgt	192
	Thr Gly Phe Ser Gly Thr Ala Ala Thr Val Leu Ile Thr Ala Lys Arg	
5	50 55 60	
	cgt gtt ttg atc aca gat tca cgt tat acc ttg ctt gct aaa gct agt	240
	Arg Val Leu Ile Thr Asp Ser Arg Tyr Thr Leu Leu Ala Lys Ala Ser	
	65 70 75 80	
10	gtt gag gga ttt gat att atc gaa agc cgc acg ccg ctt aag gtt gtg	288
	Val Glu Gly Phe Asp Ile Ile Glu Ser Arg Thr Pro Leu Lys Val Val	
	85 90 95	
15	gca gaa ttg tta gag gct gat caa ata gat tgc ctt ggt ttt gag gac	336
	Ala Glu Leu Ser Phe Glu Ala Asp Gln Ile Asp Cys Leu Gly Phe Glu Asp	
	100 105 110	
20	cag gta tcg ttt tct ttt tac cag gcc atg caa gca gaa ctg tca gga	384
	Gln Val Ser Phe Ser Phe Tyr Gln Ala Met Gln Ala Glu Leu Ser Gly	
	115 120 125	
25	ata acc ttg ctt gct cag tca ggt ttt gtg gag cat tta cgt ctt att	432
	Ile Thr Leu Leu Ala Gln Ser Gly Phe Val Glu His Leu Arg Leu Ile	
	130 135 140	
	aag gac gcc tct gaa atc gat acc att gct aaa gcg tgc tcg atc tca	480
	Lys Asp Ala Ser Glu Ile Asp Thr Ile Ala Lys Ala Cys Ser Ile Ser	
	145 150 155 160	
30	gac aaa gca ttt gaa gat gct ctt gat ttt att aaa cca ggg aca acc	528
	Asp Lys Ala Phe Glu Asp Ala Leu Asp Phe Ile Lys Pro Gly Thr Thr	
	165 170 175	
35	act gaa cgt gac ctg gct aat ttt tta gat ttt cgt atg cgt cag tat	576
	Thr Glu Arg Asp Leu Ala Asn Phe Leu Asp Phe Arg Met Arg Gln Tyr	
	180 185 190	
40	ggt gcc agc ggc aca tca ttt gat atc att gta gct tca ggc tat ctc	624
	Gly Ala Ser Gly Thr Ser Phe Asp Ile Ile Val Ala Ser Gly Tyr Leu	
	195 200 205	
45	tct gcc atg cct cat gga cgc gcc agt gac aag gtt atc cag aat aaa	672
	Ser Ala Met Pro His Gly Arg Ala Ser Asp Lys Val Ile Gln Asn Lys	
	210 215 220	
	gag agc ttg acc atg gac ttt ggg tgt tac tac aat cac tat gtt agt	720
	Glu Ser Leu Thr Met Asp Phe Gly Cys Tyr Tyr Asn His Tyr Val Ser	
	225 230 235 240	
50	gat atg acg agg acc att cat att cat att ggc caa gtt act gat gaa	768
	Asp Met Thr Arg Thr Ile His Ile His Ile Gly Gln Val Thr Asp Glu	
	245 250 255	
55	gaa cgt gag att tat gct ctt gtt ctt gct gct aat aag gct tta att	816
	Glu Arg Glu Ile Tyr Ala Leu Val Leu Ala Ala Asn Lys Ala Leu Ile	
	260 265 270	
60	gct aaa gct agc gct ggc atg act tat agt gac ttt gac ggt att ccg	864
	Ala Lys Ala Ser Ala Gly Met Thr Tyr Ser Asp Phe Asp Gly Ile Pro	
	275 280 285	
65	cgc caa ctc atc act gag gcg ggt tat ggc agt cgc ttc aca cat ggc	912
	Arg Gln Leu Ile Thr Glu Ala Gly Tyr Gly Ser Arg Phe Thr His Gly	
	290 295 300	
	att ggt cat ggc atc ggg ctt gac atc cat gag aat cca ttt ttt ggg	960
	Ile Gly His Gly Ile Gly Leu Asp Ile His Glu Asn Pro Phe Phe Gly	
	305 310 315 320	

-11-

aaa tct gag caa ctt ctc caa gct gga atg gtg gt aacagatgag 1005  
 Lys Ser Glu Gln Leu Leu Gln Ala Gly Met Val 330

5 ccagggtatct atttggataa caaatatggt gtccgtattg aagatgactt ggttatcaca 1065  
 aaactggctt gtcaagtctt gaccttggca cccaaagaat taattgtatt gtaa 1119

&lt;210&gt; 34

&lt;211&gt; 332

10 &lt;212&gt; PRT

&lt;213&gt; Streptococcus Pyogenes

&lt;400&gt; 34

15 Met Leu Gln Tyr Ser Gln Lys Leu Pro Lys Glu Phe Ala Met Ser Gly  
 1 5 10 15  
 Phe Leu Glu Gln Arg Leu Gly His Cys Leu Arg Gln Met Ala Glu Lys  
 20 25 30  
 Gly Leu Glu Ala Leu Leu Val Thr His Leu Thr Asn Ser Tyr Tyr Leu  
 35 40 45  
 20 Thr Gly Phe Ser Gly Thr Ala Ala Thr Val Leu Ile Thr Ala Lys Arg  
 50 55 60  
 Arg Val Leu Ile Thr Asp Ser Arg Tyr Thr Leu Ala Lys Ala Ser  
 65 70 75 80  
 Val Glu Gly Phe Asp Ile Ile Glu Ser Arg Thr Pro Leu Lys Val Val  
 85 90 95  
 25 Ala Glu Leu Leu Glu Ala Asp Gln Ile Asp Cys Leu Gly Phe Glu Asp  
 100 105 110  
 Gln Val Ser Phe Ser Phe Tyr Gln Ala Met Gln Ala Glu Leu Ser Gly  
 115 120 125  
 30 Ile Thr Leu Leu Ala Gln Ser Gly Phe Val Glu His Leu Arg Leu Ile  
 130 135 140  
 Lys Asp Ala Ser Glu Ile Asp Thr Ile Ala Lys Ala Cys Ser Ile Ser  
 145 150 155 160  
 Asp Lys Ala Phe Glu Asp Ala Leu Asp Phe Ile Lys Pro Gly Thr Thr  
 165 170 175  
 35 Thr Glu Arg Asp Leu Ala Asn Phe Leu Asp Phe Arg Met Arg Gln Tyr  
 180 185 190  
 Gly Ala Ser Gly Thr Ser Phe Asp Ile Ile Val Ala Ser Gly Tyr Leu  
 195 200 205  
 40 Ser Ala Met Pro His Gly Arg Ala Ser Asp Lys Val Ile Gln Asn Lys  
 210 215 220  
 Glu Ser Leu Thr Met Asp Phe Gly Cys Tyr Tyr Asn His Tyr Val Ser  
 225 230 235 240  
 Asp Met Thr Arg Thr Ile His Ile His Ile Gly Gln Val Thr Asp Glu  
 245 250 255  
 45 Glu Arg Glu Ile Tyr Ala Leu Val Leu Ala Ala Asn Lys Ala Leu Ile  
 260 265 270  
 Ala Lys Ala Ser Ala Gly Met Thr Tyr Ser Asp Phe Asp Gly Ile Pro  
 275 280 285  
 50 Arg Gln Leu Ile Thr Glu Ala Gly Tyr Gly Ser Arg Phe Thr His Gly  
 290 295 300  
 Ile Gly His Gly Ile Gly Leu Asp Ile His Glu Asn Pro Phe Phe Gly  
 305 310 315 320  
 Lys Ser Glu Gln Leu Leu Gln Ala Gly Met Val Val 330  
 55 325



# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/16229

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12N 9/10, 15/63, 1/21, 1/15, 1/19; C12Q 1/48; C07H 21/04

US CL : 435/193, 320.1, 252.3, 254.11, 15, 7.32; 536/23.2

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/193, 320.1, 252.3, 254.11, 15, 7.32; 536/23.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Database PROMT on STN, Information Access Company, AN 97:420001, 'Antibacterials, SIGA Wyeth Ayerst, SIGA licensing agreement.' abstract. R & D Focus Drug News, 28 July 1997, see whole abstract.	1-49, 57-71 and 95-115



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*B* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

02 NOVEMBER 1998

Date of mailing of the international search report

19 NOV 1998

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

KAWAI LAU

Telephone No. (703) 308-0196

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/16229

## Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
1-49, 57-71 and 95-115

Remark on Protest

☐  
☐

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/16229

### B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, STN: PROMT, CAPLUS, CEN, MEDLINE, BIOSIS, BIOTECHDS, CANCERLIT, CEABA, DRUGNL,  
EMBASE, FSTA, SCISEARCH, TOXLINE, PIRA, WPIDS, EUROPATFULL

search terms: sortase, sort, enzyme, transamidase, gram, positive

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☒ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**